NASA Technical Memorandum 105608
ICOMP–92–09

Upper Bounds for Convergence Rates of Vector Extrapolation Methods on Linear Systems With Initial Iterations

Avram Sidi
Technion–Israel Institute of Technology
Haifa, Israel

and Institute for Computational Mechanics in Propulsion
Lewis Research Center
Cleveland, Ohio

and

Yair Shapira
Technion–Israel Institute of Technology
Haifa, Israel

July 1992

NASA

Upper Bounds for Convergence Rates of Vector Extrapolation Methods on Linear Systems with Initial Iterations

Avram Sidi
Computer Science Department
Technion - Israel Institute of Technology
Haifa 32000, Israel
and
Institute for Computational Mechanics in Propulsion
NASA - Lewis Research Center
Cleveland, Ohio 44135, U.S.A.
and
Yair Shapira
Mathematics Department
Technion - Israel Institute of Technology
Haifa 32000, Israel

Abstract

The application of the minimal polynomial extrapolation (MPE) and the reduced rank extrapolation (RRE) to a vector sequence obtained by the linear iterative technique $x_{j+1} = Ax_j + b,\ j = 1, 2, ...,$ is considered. Both methods produce a two-dimensional array of approximations $s_{n,k}$ to the solution of the system $(I - A)x = b$. Here $s_{n,k}$ is obtained from the vectors $x_j,\ n \leq j \leq n + k + 1$. It was observed in an earlier publication by the first author that the sequence $s_{n,k},\ k = 1, 2, ...,$ for $n > 0$, but fixed, possesses better convergence properties than the sequence $s_{0,k},\ k = 1, 2, ...$. A detailed theoretical explanation for this phenomenon is provided in the present work. This explanation is heavily based on approximations by incomplete polynomials. It is demonstrated by numerical examples when the matrix A is sparse that cycling with $s_{n,k}$ for $n > 0$, but fixed, produces better convergence rates and costs less computationally than cycling with $s_{0,k}$. It is also illustrated numerically with a convection-diffusion problem that the former may produce excellent results where the latter may fail completely. As has been shown in an earlier publication, the results produced by $s_{0,k}$ are identical to the corresponding results obtained by applying the Arnoldi method or GMRES to the system $(I - A)x = b$.

1 Introduction

Let s be the solution to the nonsingular linear $N \times N$ system $Bx = f$, which we choose to write equivalently in the possible preconditioned form

$$x = Ax + b. \tag{1.1}$$

With x_0 picked arbitrarily, we iterate (1.1), and generate the vectors $x_1, x_2, \ldots$, i.e.,

$$x_{j+1} = Ax_j + b, \quad j = 0, 1, \ldots . \tag{1.2}$$

If $r \equiv \rho(A)$, the spectral radius of A, then the error $x_n - s$ tends to zero practically as r^n for $n \to \infty$ provided $r < 1$.

In most cases of interest that occur in practice r may be very close to 1, and this causes the sequence $\{x_j\}_{j=0}^{\infty}$ to converge very slowly. One efficient way of overcoming this problem is to use vector extrapolation methods in conjunction with the iterative scheme in (1.2). Of the various extrapolation methods two have proved to be especially effective, and these are the minimal polynomial extrapolation (MPE) of Cabay and Jackson [CaJa] and the reduced rank extrapolation (RRE) of Eddy [Ed] and Mešina [M]. A method almost identical to RRE was given earlier by Kaniel and Stein [KStei].

Both MPE and RRE, when applied to the sequence $\{x_j\}_{j=0}^{\infty}$, produce a two-dimensional array of approximations to s, whose entries we denote by $s_{n,k}$. For given integers $n \geq 0$ and $k \geq 1$, $s_{n,k}$ for both methods is determined from the vectors $x_n, x_{n+1}, \ldots, x_{n+k+1}$, and is of the form

$$s_{n,k} = \sum_{j=0}^{k} \gamma_j^{(n,k)} x_{n+j}, \tag{1.3}$$

such that the scalars $\gamma_j^{(n,k)}$ depend on x_i, $n \leq i \leq n+k+1$, nonlinearly, and satisfy

$$\sum_{j=0}^{k} \gamma_j^{(n,k)} = 1. \tag{1.4}$$

Thus the indices n and k of $s_{n,k}$ indicate that n iterations have been performed with the iterative scheme in (1.2) and that MPE or RRE is being applied to $x_n, x_{n+1}, \ldots, x_{n+k+1}$, the initial vectors $x_0, x_1, \ldots, x_{n-1}$ being discarded.

A discussion of MPE and RRE in the setting of both linear and nonlinear iterative techniques can be found in the survey paper of Smith, Ford, and Sidi [SmFoSi], where other related literature is also cited. For a different approach see the paper by Sidi, Ford, and Smith [SiFoSm].

Convergence properties of MPE and RRE for sequences generated by a scheme such as (1.2) have been studied extensively in various papers. The convergence of $s_{n,k}$ for $n \to \infty$ is the subject of a paper by Sidi [Si1] and of another by Sidi and Bridger [SiB]. The first of these papers is concerned with the case in which the matrix A in (1.2) is diagonalizable, while the second is concerned with a defective matrix A. The behavior of $s_{n,k}$ for fixed n and increasing k is the topic of an additional paper by Sidi [Si2]. This paper also discusses the equivalence of MPE and RRE and other related vector extrapolation methods, as they are applied in conjunction with (1.2), with Krylov subspace methods, as these are applied to the linear system $(I-A)x = b$. In particular, it is shown in [Si2] that the vectors $s_{0,k}$ generated by MPE and RRE are precisely those generated by the method of Arnoldi [Ar] (see also Saad [Saa]) and the method of generalized conjugate residuals (GCR) of Eisenstat, Elman, and Schultz [EiElSc], respectively. (In other words, Krylov subspace methods produce only the first rows of the arrays of approximations produced by the corresponding extrapolation methods.) The conjugate gradient type method of Axelson [Ax], the method of Young and Jea [YJe] that has been designated ORTHODIR, and the generalized minimum residual method (GMRES) of Saad and Schultz [SaaSc] are all mathematically equivalent to GCR. In case the matrix $I - A$ is hermitian, the method of Arnoldi and GCR are equivalent to the method of conjugate gradients of Hestenes and Stiefel [HSti] and the method of conjugate residuals of Stiefel [Sti], respectively, and when A is antihermitian, they are equivalent to the method of generalized conjugate gradients (GCG) of Concus and Golub [CoGo] and Widlund [W] and to ORTHOMIN(1) of Vinsome [Vi], respectively. It should be mentioned, though, that, unlike Krylov subspace methods that can be used in the solution of linear systems only, MPE and RRE and other vector extrapolation methods can be employed in the solution of nonlinear as well as linear systems. The reason for this is that vector extrapolation methods are defined in terms of a vector sequence, and whether this sequence arises from iterative solution of a linear or nonlinear system is irrelevant. The Krylov subspace methods, on the other hand, make direct use of the matrix of the linear system being solved. In addition, they are not based on any fixed point iterative method for this linear system.

In Ford and Sidi [FoSi] the existence of an interesting four-term recursion relation among the $s_{n,k}$ is shown. This recursion relation is of the form

$$s_{n,k+1} = \alpha_{n,k} s_{n,k} + \beta_{n,k} s_{n+1,k-1} + (1 - \alpha_{n,k} - \beta_{n,k}) s_{n+1,k} \tag{1.5}$$

for some scalars $\alpha_{n,k}$ and $\beta_{n,k}$.

Finally, in a recent work by Sidi [Si3] efficient and numerically stable implementations of MPE

and RRE are given. This work also contains a FORTRAN 77 program that was used in producing the numerical results reported in Section 6 of the present work.

As mentioned in [Si3], when one applies MPE and RRE to a linear system in the so called cycling mode (to be described later in this section) with $s_{n,k}$, n and k being held fixed, much better convergence behavior is observed for even moderate values of n than for $n = 0$. It may also happen that no noticeable convergence takes place for $n = 0$. This is a very curious phenomenon, which we would like to try to explain in this work. It is obvious that any explanation of it would have to be through the analysis of $s_{n,k} - s$ for *finite* values of n and k.

We would like to emphasize that the results of [Si1] and [SiB] concerning $s_{n,k} - s$ are *asymptotic* in nature, i.e., they capture the true behavior of $s_{n,k}$ for $n \to \infty$ with k fixed, in an optimal way. For example, if we assume the matrix A to be diagonalizable, and order its distinct nonzero eigenvalues $\lambda_1, \lambda_2, ...$, such that $|\lambda_1| \geq |\lambda_2| \geq \cdots$, then, provided $|\lambda_k| > |\lambda_{k+1}|$ and $x_0 - s$ has contributions from each of the invariant subspaces of A associated with $\lambda_1, \lambda_2, ..., \lambda_k$, we have, for both MPE and RRE,

$$s_{n,k} - s = O(|\lambda_{k+1}|^n) \quad \text{as } n \to \infty. \tag{1.6}$$

Naturally, a result such as this, although good for sufficiently large n, cannot explain the behavior of $s_{n,k}$ for small or moderately large values of n and arbitrary values of k.

The results of [Si2], on the other hand, are stated in terms of inequalities, hence might be considered appropriate for all values of n and k. For example, when the matrix A is diagonalizable, and all the eigenvalues of $I - A$ are real and positive, $\mu_{\max}$ and $\mu_{\min}$ being, respectively, the largest and smallest of these eigenvalues, for both MPE and RRE,

$$||s_{0,k} - s|| \leq K\eta^k ||x_0 - s||, \tag{1.7}$$

where

$$\eta = \frac{\sqrt{\kappa} - 1}{\sqrt{\kappa} + 1}, \quad \kappa \equiv \frac{\mu_{\max}}{\mu_{\min}}, \tag{1.8}$$

and $||\cdot||$ is any vector norm and K is an appropriate positive constant independent of k. This result basically tells that $||s_{0,k} - s||$ is smaller than $||x_0 - s||$ practically by a factor of η^k. Recalling now that $s_{n,k}$ is the result of applying MPE or RRE to the vector sequence $\{x_j\}_{j=0}^{\infty}$ starting with x_n instead of x_0, for all n and k, we can replace (1.7) by

$$||s_{n,k} - s|| \leq K\eta^k ||x_n - s||. \tag{1.9}$$

This inequality provides us with an upper bound on $||s_{n,k} - s||/||x_n - s||$. Ultimately, however, we would like to have an upper bound on $||s_{n,k} - s||/||x_0 - s||$. For this we make use of

$$x_n - s = A^n(x_0 - s) \tag{1.10}$$

in 1.9), obtaining finally

$$||s_{n,k} - s|| \leq K'\eta^k r^n ||x_0 - s||, \tag{1.11}$$

where K' is an appropriate positive constant independent of k, and $r = \rho(A)$, as before.

Despite the fact that (1.9) and (1.11) hold for all n and k when A is diagonalizable and the eigenvalues of $I - A$ are real and positive, both of these inequalities are much too crude and pessimistic. In addition, as has been mentioned in [Si3], and has been observed in many numerical examples, the convergence of $s_{n,k}$ to s is far better than suggested by both (1.9) and (1.11) when n is even moderately large. In other words, for increasing k, better convergence rates are observed for $s_{n,k}$ with $n > 0$ than for $s_{0,k}$. Actually, this should be expected judging from the symptotic result in (1.6), although the latter cannot be used to quantify this interesting phenomenon.

In addition to being interesting, the above mentioned phenomenon is also potentially very useful in the following sense: Suppose MPE or RRE is being applied in conjunction with the iterative scheme of (1.2) in the cycling mode. This is achieved by performing the steps below.

$$\left\{\begin{array}{ll} \text{Step 0.} & \text{Pick } x_0, n, \text{ and } k. \\ \text{Step 1.} & \text{Compute } x_1, x_2, ..., x_n, ..., x_{n+k+1} \text{ by (1.2),} \\ & \text{and compute } s_{n,k}. \\ \text{Step 2.} & \text{If } s_{n,k} \text{ satisfies accuracy test, stop;} \\ & \text{otherwise set } x_0 = s_{n,k}, \text{ and go to Step 1.} \end{array}\right. \tag{1.12}$$

Doing Step 1 once is called a cycle. Now in the cycling mode the accuracy test may be passed by $s_{n,k}$ with $n > 0$ in fewer cycles than by $s_{0,k}$. Even though the number of iterations with (1.2) is $n + k + 1$ for $s_{n,k}$ as opposed to only $k + 1$ for $s_{0,k}$ in each cycle, the overhead caused by the application of MPE or RRE may increase significantly the computational cost of cycling with $s_{0,k}$. We may thus end up paying a higher price for cycling with $s_{0,k}$ than with $s_{n,k}$ for $n > 0$. This will be especially pronounced in cases where the iterative scheme in (1.2) is inexpensive, which may come about if A is very sparse. It may also happen that the total number of iterations with (1.2) will be less for cycling with $s_{n,k}$, $n > 0$, than for cycling with $s_{0,k}$. It may even occur that cycling with $s_{0,k}$ stalls numerically, whereas cycling with $s_{n,k}$, even with a moderate value of n, can produce very quick convergence. All this has been observed in many numerical examples.

Our purpose in the present work is to provide a rigorous explanation of this phenomenon. We accomplish this by deriving upper bounds for $||s_{n,k} - s||$ of the form

$$||s_{n,k} - s|| \leq f(n,k)||x_0 - s||, \tag{1.13}$$

which capture the true behavior of $s_{n,k}$ quite accurately. This is done in Section 3. In Section 4 we derive some easily computable upper bounds $f(n,k)$ for certain cases. This is accomplished through approximate solutions to some best approximation problems by incomplete polynomials, whose near-best qualities are verified numerically in Section 5. Finally, in Section 6 we give two numerical examples to support all the claims that we make throughout the paper. One of these examples involves the application of MPE or RRE to a linear system that arises from a finite difference discretization of a two-dimensional convection-diffusion equation. The convergence of $s_{n,k}$ for moderately large n is extremely quick in the cycling mode even when the underlying iterative scheme is divergent, whereas $s_{0,k}$ in the cycling mode stalls or is very slow at best. Cycling with $s_{0,k}$ using RRE produces results identical to those that are obtained from GCR(k) or its equivalent GMRES(k), as follows from [Si2] and as mentioned in the previous section.

The examples of Section 6 and the theory given in Sections 3 and 4 thus make it clear that vector extrapolation methods may be more flexible and may achieve better accuracy than Krylov subspace methods, and may produce very good results also where the latter may fail completely.

2 Technical Preliminaries

In the previous section we mentioned that the approximations $s_{n,k}$ to s, obtained from the vector sequence $\{x_j\}_{j=0}^{\infty}$ are of the form given in (1.3) and (1.4). For MPE the scalars $\gamma_j^{(n,k)}$ are defined by the linear equations

$$\sum_{j=0}^{k}(u_{n+i}, u_{n+j})\gamma_j^{(n,k)} = 0, \quad 0 \leq i \leq k-1,$$

$$\sum_{j=0}^{k}\gamma_j^{(n,k)} = 1, \tag{2.1}$$

whereas for RRE they are defined by

$$\sum_{j=0}^{k}(w_{n+i}, u_{n+j})\gamma_j^{(n,k)} = 0, \quad 0 \leq i \leq k-1,$$

$$\sum_{j=0}^{k}\gamma_j^{(n,k)} = 1, \tag{2.2}$$

where

$$u_i = \Delta x_i = x_{i+1} - x_i \text{ and } w_i = \Delta u_i = \Delta^2 x_i, \quad i = 0, 1, ..., \tag{2.3}$$

and $(\cdot,\cdot)$ is the Euclidean inner product on $\mathbf{C}^N$. These equations and ensuing determinantal representations for $s_{n,k}$ were first presented in [Si1]. The determinantal representations were very useful in the convergence analysis of $s_{n,k}$ for $n \to \infty$ with k fixed.

When the vector sequence $\{x_j\}_{j=0}^{\infty}$ is generated by the linear iterative scheme in (1.2), we define the residual vector $r(x)$ associated with an arbitrary vector x by

$$r(x) = Ax + b - x. \tag{2.4}$$

We also define the matrix C and its hermitian part C_{H} by

$$C = I - A \text{ and } C_{\mathrm{H}} = \frac{1}{2}(C + C^*). \tag{2.5}$$

We let $||\cdot||$ denote the vector l_2-norm induced by the Euclidean inner product in $\mathbf{C}^N$, or the operator norm induced by this vector norm. In addition, in case C_{H} is positive definite, we define the vector norm $||\cdot||'$ by

$$||x||' = \sqrt{(x, C_{\mathrm{H}} x)}, \tag{2.6}$$

and let $||\cdot||'$ stand for the induced operator norm as well.

It is shown that when the vector sequence $\{x_j\}_{j=0}^{\infty}$ is generated by a linear iterative scheme such as (1.2), $s_{n,k}$, for both MPE and RRE, exists and is equal to s, provided k is the degree of the minimal polynomial of the matrix A with respect to the vector $u_n = \Delta x_n = x_{n+1} - x_n$. It has been shown in [Si2] that when k is less than this degree, then $s_{n,k}$ for RRE always exists, but $s_{n,k}$ for MPE does not necessarily exist. A sufficient condition for existence of $s_{n,k}$ for MPE in this case is that C_{H} be positive definite, see [Si2].

We now state a result concerning the error $s_{n,k} - s$ that has been given in [Si2].

Theorem 2.1. *For RRE*

$$||r(s_{n,k})|| \leq ||Q_k(C) r(x_n)||, \tag{2.7}$$

while for MPE, assuming that C_{H} is positive definite,

$$||s_{n,k} - s||' \leq L ||Q_k(C)(x_n - s)||', \tag{2.8}$$

where L is a constant given as

$$L = ||C_{\mathrm{H}}^{-\frac{1}{2}} C C_{\mathrm{H}}^{-\frac{1}{2}}||. \tag{2.9}$$

In both (2.7) and (2.8), $Q_k(z)$ is an arbitrary polynomial of degree at most k that satisfies $Q_k(0) = 1$.

The result in (2.7) follows from the analysis of GCR given in [EiElSc] and the equivalence of RRE and GCR that is proved in [Si2]. In [Si2] a unified approach is presented from which both (2.7) and (2.8) can be obtainted simultaneously. Theorem 2.1 will be the starting point of our analysis in the next section.

Before closing this section we mention that a result such as (1.9) can be obtained from Theorem 2.1 by replacing the right hand sides of (2.7) and (2.8) by

$$||r(s_{n,k})|| \leq ||Q_k(C)||\ ||r(x_n)|| \quad \text{for RRE} \tag{2.10}$$

and

$$||s_{n,k} - s||' \leq L||Q_k(C)||'\ ||x_n - s||' \quad \text{for MPE,} \tag{2.11}$$

respectively. For further details and developments, see [Si2].

3 Derivation of Upper Bounds

Theorem 3.1 is one of the main results of this section. We use the notation of Sections 1 and 2 throughout.

Theorem 3.1. *Define*

$$\Gamma_{n,k} = \min_{Q_k} ||A^n Q_k(C)|| \tag{3.1}$$

and

$$\Gamma'_{n,k} = \min_{Q_k} ||A^n Q_k(C)||', \tag{3.2}$$

where $Q_k(z)$ are polynomials of degree at most k that satisfy $Q_k(0) = 1$. Then

$$||r(s_{n,k})|| \leq \Gamma_{n,k}||r(x_0)|| \quad \textit{for RRE} \tag{3.3}$$

and, provided C_H is positive definite,

$$||s_{n,k} - s||' \leq L\Gamma'_{n,k}||x_0 - s||' \quad \textit{for MPE,} \tag{3.4}$$

with L as given in (2.9).

Proof. First we note that

$$r(x) = C(s - x) \tag{3.5}$$

and

$$x_n - s = A^n(x_0 - s), \text{ for all } n. \tag{3.6}$$

Substituting (3.6) in (2.7) first, and using (3.5) next, we obtain

$$||r(s_{n,k})|| \leq ||Q_k(C)A^n r(x_0)|| \text{ for RRE.} \tag{3.7}$$

Here we have also made use of the fact that A and C commute, which is a result of (2.5). The result in (3.3) now follows from (3.7) if we also recall that the polynomial $Q_k(z)$ in Theorem 2.1 is of degree at most k and satisfies $Q_k(0) = 1$, but is arbitrary otherwise. The result in (3.4) can be obtained from (2.8) in exactly the same way. □

As has been shown in [Si2], for any matrix G,

$$||G||' = ||C_{\mathrm{H}}^{\frac{1}{2}} G C_{\mathrm{E}}^{-\frac{1}{2}}||, \tag{3.8}$$

from which we have

$$\begin{aligned} ||G||' &= ||G|| \quad \text{if } GC_{\mathrm{H}} = C_{\mathrm{H}}G, \\ ||G||' &\leq \sqrt{\mathrm{cond}_2(C_{\mathrm{E}})}||G|| \quad \text{otherwise.} \end{aligned} \tag{3.9}$$

Using (3.9) in (3.2), we obtain

$$\begin{aligned} \Gamma'_{n,k} &= \Gamma_{n,k} \quad \text{if } C \text{ normal,} \\ \Gamma'_{n,k} &\leq \sqrt{\mathrm{cond}_2(C_{\mathrm{E}})}\Gamma_{n,k} \quad \text{otherwise.} \end{aligned} \tag{3.10}$$

This result enables us to unify the treatments of MPE and RRE, as $\Gamma_{n,k}$ is now the only important quantity that needs to be analyzed as a function of n and k.

It is very instructive to compare, e.g., the two bounds concerning $s_{n,k}$ for RRE that are given in (2.10) and (3.3). We observe that the matrix A^n in (2.10) forms part of $||r(x_n)||$, whereas it is part of the operator $A^nQ_k(C)$ in (3.3). It is this shift in the location of A^n that makes the difference between the qualities of these two upper bounds.

Hereafter we assume for simplicity that the matrix A is diagonalizable. We shall denote the eigenvalues of A by $\lambda_1, \lambda_2 \ldots, \lambda_N$, and the matrix that diagonalizes A by R, so that

$$A = R\,\Lambda\,R^{-1}, \ \Lambda = \ \mathrm{diag}\,(\lambda_1, \lambda_2, ..., \lambda_N). \tag{3.11}$$

We shall also define

$$\mathcal{P}_k = \{q(\lambda) = \sum_{i=0}^{k} a_i \lambda^i : q(1) = 1\}. \tag{3.12}$$

Theorem 3.2. *Define*

$$\Gamma^*_{n,k} = \min_{q \in \mathcal{P}_k} \max_i |\lambda_i^n q(\lambda_i)|. \tag{3.13}$$

Then

$$\Gamma_{n,k} = \Gamma'_{n,k} = \Gamma^*_{n,k} \quad \textit{if } C \textit{ (or } A\textit{) normal} \tag{3.14}$$

and

$$\begin{aligned} \Gamma_{n,k} &\leq \mathrm{cond}_2(R)\Gamma^*_{n,k} \\ &\qquad\qquad \textit{if } C \textit{ (or } A\textit{) nonnormal.} \\ \Gamma'_{n,k} &\leq \mathrm{cond}_2(C_H^{\frac{1}{2}} R)\Gamma^*_{n,k} \end{aligned} \tag{3.15}$$

Proof. The results above follow by substituting (2.5) and (3.11) in (3.1) and (3.2), and realizing that the matrix $\Lambda^n q(\Lambda)$ is diagonal. We also need to use (3.8) with $\Gamma'_{n,k}$. The details are left to the reader. □

Note: It is obvious that for $n > 0$ the maximum in (3.13) is being considered on the nonzero λ_i only.

We can now combine Theorems 3.1 and 3.2, and obtain bounds in terms of $\Gamma^*_{n,k}$. For the sake of completeness, these are summarized below as Theorem 3.3.

Theorem 3.3. *Under the conditions of Theorem 3.2, for RRE*

$$\frac{||r(s_{n,k})||}{||r(x_0)||} \leq \begin{cases} \Gamma^*_{n,k} & \textit{if } C \textit{ (or } A\textit{) normal,} \\ \mathrm{cond}_2(R)\Gamma^*_{n,k} & \textit{otherwise,} \end{cases} \tag{3.16}$$

and for MPE

$$\frac{||s_{n,k} - s||'}{||x_0 - s||'} \leq \begin{cases} L\Gamma^*_{n,k} & \textit{if } C \textit{ (or } A\textit{) normal,} \\ L\mathrm{cond}_2(C_H^{\frac{1}{2}} R)\Gamma^*_{n,k} & \textit{otherwise,} \end{cases} \tag{3.17}$$

with L as given in (2.9).

It turns out that the upper bounds given by (3.16) and (3.17) are quite tight when A is normal or when $\mathrm{cond}_2(R)$ is small. When $\mathrm{cond}_2(R)$ is large, however, they become pessimistic. In this case one has to go back to (3.3) and (3.4) which are still good, and try to bound $\Gamma_{n,k}$ and $\Gamma'_{n,k}$ in a manner different from Theorem 3.2.

Finally, by applying Theorem 3.1 to cycling we obtain the following result.

Theorem 3.4. *Let $n \geq 0$ and $k > 0$ be fixed integers and denote by $s_{n,k}^{(i)}$ the approximation $s_{n,k}$ that is obtained at the end of the ith cycle when MPE or RRE is being used in the cycling mode as described in (1.12). Then, under the conditions of Theorem 3.1,*

$$||r(s_{n,k}^{(i)})|| \leq (\Gamma_{n,k})^i ||r(x_{\text{init}})|| \quad \textit{for RRE} \tag{3.18}$$

and

$$||s_{n,k}^{(i)} - s||' \leq (L\Gamma'_{n,k})^i ||x_{\text{init}} - s||' \quad \textit{for MPE}, \tag{3.19}$$

where x_{init} is the initial vector at the beginning of the first cycle.

4 Bounds on $\Gamma^*_{n,k}$

From the definition of $\Gamma^*_{n,k}$ given in (3.13) it is obvious that precise knowledge of it requires complete information on the spectrum of A, which is not available in general. We may, however, obtain very good bounds on $\Gamma^*_{n,k}$ for $n > 0$ if we know that the nonzero part of the spectrum of A is contained in a set D of the complex λ-plane that does not contain 1. (If 1 is an eigenvalue of A, the system in (1.1) is singular, contrary to our assumption in the introduction.) Then, with $\mathcal{P}_k$ as in (3.12),

$$\Gamma^*_{n,k} \leq \min_{q \in \mathcal{P}_k} \max_{\lambda \in D} |\lambda^n q(\lambda)| \equiv \Gamma^D_{n,k}. \tag{4.1}$$

If D is a domain, then, by the maximum modulus theorem for analytic functions,

$$\Gamma^*_{n,k} \leq \Gamma^D_{n,k} = \min_{q \in \mathcal{P}_k} \max_{\lambda \in \partial D} |\lambda^n q(\lambda)|, \tag{4.2}$$

where ∂D denotes the boundary of D.

In general, the min-max problems of (4.1) and (4.2) cannot be solved analytically in a simple way. When $n = 0$, some analytic solutions are known, however. The best known is the one for the case in which D is a finite real interval $[\alpha, \beta]$ with $\alpha < \beta < 1$, α being arbitrary otherwise. In this case the optimal polynomial $q(\lambda)$ is $T_k(g(\lambda))/T_k(g(1))$, with $g(\lambda) = (2\lambda - \alpha - \beta)/(\beta - \alpha)$, where $T_k(z)$ is the Chebyshev polynomial of the first kind of degree k. Consequently,

$$\Gamma^D_{0,k} = \frac{1}{T_k(\frac{2-\alpha-\beta}{\beta-\alpha})}. \tag{4.3}$$

This result can be found in, e.g., Varga [Va]. We mention in passing that the result in (1.9) can be obtained directly from (4.3). The analytical solution of the min-max problem for $D = \{\lambda : \lambda = i\xi, -\beta \leq \xi \leq \beta, \beta > 0 \text{ real}\}$ has been provided recently by Freund and Ruscheweyh [FrR], who also

give a numerical method for the case in which D is any line segment in the complex λ-plane not containing 1.

We are not aware of any solution to the min-max problem in (4.1) that is known analytically when $n > 0$. Instead of trying to solve this problem, we shall derive easily computable upper bounds for $\Gamma_{n,k}^D$ for all n and k for some sets D.

4.1. *The Case $D = [\alpha, \beta]$, $0 < \alpha < \beta < 1$*

This is the simplest and most instructive case and we would like to analyze it in some detail. The result that we obtain for this case will eventually show the way to obtain good bounds for $\Gamma_{n,k}^D$ for other sets D as well.

We start by observing that the min-max problems in (4.1) and (4.2) are constrained best uniform approximation problems by incomplete polynomials. The problem relevant to the present case is the one in (4.1), which now reads

$$\Gamma_{n,k}^D = \min_{\sum_{i=0}^{k} a_i = 1} \max_{\alpha \le \lambda \le \beta} \left| \sum_{i=0}^{k} a_i \lambda^{n+i} \right|. \tag{4.4}$$

Uniform approximation on the real interval [0,1] by incomplete polynomials has been studied by Lorentz [L] and, in a series of papers, by Saff and Varga [Saf Va1, Saf Va2, Saf Va3, Saf Va4].

The following result is similar to a corresponding result in [SafVa1] that was proved for the interval [0,1].

Lemma 4.1. *There exists a unique monic polynomial $p^*(\lambda)$ of degree k that satisfies*

$$\max_{\alpha \le \lambda \le \beta} |\lambda^n p^*(\lambda)| \le \max_{\alpha \le \lambda \le \beta} |\lambda^n p(\lambda)|, \tag{4.5}$$

where $p(\lambda)$ is any monic polynomial of degree k. Also there exist $k+1$ points $t_1 < t_2 < \cdots < t_{k+1}$, all in $[\alpha, \beta]$, at which $\lambda^n p^*(\lambda)$ takes on the value $\max_{\alpha \le \lambda \le \beta} |\lambda^n p^*(\lambda)|$ with alternating signs.

Proof. The assertion above follows from the fact the functions $\lambda^n, \lambda^{n+1}, \ldots, \lambda^{n+k-1}$ form a k-dimensional Haar subspace on $[\alpha, \beta]$ since $\alpha > 0$. Then $\lambda^n p(\lambda)$ is simply the error incurred by approximating the function λ^{n+k} by some function in this Haar subspace. The proof is now completed by employing the uniqueness theorem and the alternation theprem on p.80 and p.75, respectively, in Cheney [Ch]. We leave the details to the reader. □

Theorem 4.2. *Let $p^*(\lambda)$ be as in Lemma 4.1. Then the polynomial $q^*(\lambda)$, which is given by*

$$q^*(\lambda) = \frac{p^*(\lambda)}{p^*(1)}, \tag{4.6}$$

solves the min-max problem in (4.4). Consequently,

$$\Gamma^D_{n,k} = \frac{\max_{\alpha\le\lambda\le\beta} |\lambda^n p^*(\lambda)|}{|p^*(1)|}. \tag{4.7}$$

Proof. Suppose there is another polynomial $q(\lambda) \in \mathcal{P}_k$ that solves the min-max problem. Then, necessarily,

$$\max_{\alpha\le\lambda\le\beta} |\lambda^n q(\lambda)| \le \max_{\alpha\le\lambda\le\beta} |\lambda^n q^*(\lambda)|,$$

which implies that the the sign of $F(\lambda) = \lambda^n q^*(\lambda) - \lambda^n q(\lambda)$ at the points t_i of the previous lemma is that of $\lambda^n q^*(\lambda)$ at the same points. Consequently, $F(\lambda)$ has at least k zeros in $[\alpha,\beta]$. In addition, $F(1) = 0$ by the fact that $q(1) = q^*(1) = 1$. Since 1 is not in $[\alpha,\beta]$, we see that $F(\lambda)$ vanishes at least at $k+1$ points in $[\alpha,1]$. But $F(\lambda)$ is in the $(k+1)$-dimensional Haar subspace on $[\alpha,1]$ spanned by the functions $\lambda^n, \lambda^{n+1}, ..., \lambda^{n+k}$, and, therefore, can vanish at most k times on $[\alpha,1]$. We have thus a contradiction. Therefore, $q^*(\lambda)$ in (4.6) is the solution to the min-max problem. The proof of (4.7) is now immediate. □

We shall not attempt to determine $p^*(\lambda)$ analytically. We could determine $p^*(\lambda)$ numerically by the Remes algorithm, see [Ch], although this would not provide us with an analytical upper bound for $\Gamma^*_{n,k}$. Instead of doing this we shall try to give an analytical upper bound on $\Gamma^D_{n,k}$ in terms of orthogonal polynomials. If we let $\phi_{n,k}(\lambda)$ be the monic orthogonal polynomial of degree k with respect to the weight function λ^{2n} on $[\alpha,\beta]$, then we can write

$$\Gamma^D_{n,k} \le \frac{\max_{\alpha\le\lambda\le\beta} |\lambda^n \phi_{n,k}(\lambda)|}{|\phi_{n,k}(1)|}. \tag{4.8}$$

Next, by employing Theorem A.3 from the Appendix, we have

$$\max_{\alpha\le\lambda\le\beta} |\lambda^n \phi_{n,k}(\lambda)| = \beta^n |\phi_{n,k}(\beta)|, \tag{4.9}$$

so that (4.8) becomes simply

$$\Gamma^D_{n,k} \le \beta^n \frac{|\phi_{n,k}(\beta)|}{|\phi_{n,k}(1)|} \tag{4.10}$$

We now give a heuristic argument to justify the replacement of $p^*(\lambda)$ by $\phi_{n,k}(\lambda)$. First, we note that this has come about by the replacement of the best l_∞-approximation problem

$$\min_p \max_{\alpha\le\lambda\le\beta} |\lambda^n p(\lambda)|, \quad p(\lambda) \text{ monic of degree } k, \tag{4.11}$$

by the best l_2-approximation problem

$$\min_p \left\{ \int_\alpha^\beta |\lambda^n p(\lambda)|^2 d\lambda \right\}^{\frac{1}{2}}, \quad p(\lambda) \text{ monic of degree } k, \tag{4.12}$$

$p(\lambda) = p^*(\lambda)$ and $p(\lambda) = \phi_{n,k}(\lambda)$ being the solutions to the former and the latter problems, respectively. Next, especially for large values of n, $\phi_{n,k}(\lambda)$ has the most important characteristics of $p^*(\lambda)$: (i) As follows from Lemma 4.1, $p^*(\lambda)$ has precisely k simple zeros in (α, β). Being the orthogonal polynomial of degree k on $[\alpha, \beta]$, $\phi_{n,k}(\lambda)$ too has precisely k simple zeros on (α, β). (ii) Since for large n the weight functions λ^n in (4.11) and λ^{2n} in (4.12) are much more pronounced in a neighbourhood of β, the zeros of both $p^*(\lambda)$ and $\phi_{n,k}(\lambda)$ will tend to be in a neighbourhood of β as well.

The quality of the upper bound in (4.10) can be assessed by comparing it with the lower bound that follows from Theorem A.4 in the appendix. This is summarized in Theorem 4.3 below.

Theorem 4.3. $\Gamma^D_{n,k}$ *satisfies the inequalities*

$$\{(\beta - \alpha) \sum_{j=0}^{k} |\phi_{n,j}(1)|^2 / \nu_j\}^{-\frac{1}{2}} \leq \Gamma^D_{n,k} \leq \beta^n \frac{|\phi_{n,k}(\beta)|}{|\varphi_{n,k}(1)|}, \tag{4.13}$$

where

$$\nu_j = \int_\alpha^\beta \lambda^{2n} |\varphi_{n,j}(\lambda)|^2 d\lambda, \quad j = 0, 1, \ldots \ . \tag{4.14}$$

Unfortunately, the polynomials $\varphi_{n,k}(\lambda)$ are not available analytically, hence analytical forms for the upper and lower bounds on $\Gamma^D_{n,k}$ are not known for this case. The source of this problem is the fact that $\alpha > 0$. Interestingly enough, if α is replaced by zero, then $\phi_{n,k}(\lambda)$ is expressible in terms of Jacobi polynomials. In fact, $\varphi_{n,k}(\lambda)$, which now is the kth orthogonal polynomial with respect to the weight function λ^{2n} on $[0, \beta]$, is a constant multiple of $P_k^{(0,2n)}(2\lambda/\beta - 1)$ by (A.4) in the appendix. First, we observe that

$$\Gamma^D_{n,k} \leq \min_{q \in \mathcal{P}_k} \max_{0 \leq \lambda \leq \beta} |\lambda^n q(\lambda)| \equiv \Gamma^{D'}_{n,k}, \ D' = [0, \beta], \tag{4.15}$$

and that, for large n, $\Gamma^{D'}_{n,k}$ will not be too different from $\Gamma^D_{n,k}$. The reason for this is that the weight λ^n in the interval $[0, \alpha]$ is negligible compared to its average value in the interval $[\alpha, \beta]$, hence there cannot be a great difference between the solutions of the two min-max problems on $[\alpha, \beta]$ and on $[0, \beta]$. Next, by (A.2) and (A.3) in the appendix, we have, respectively,

$$\int_0^\beta \lambda^{2n} \left[P_k^{(0,2n)}(2\lambda/\beta - 1) \right]^2 d\lambda = \frac{\beta^{2n+1}}{2n + 2k + 1}, \tag{4.16}$$

and

$$P_k^{(0,2n)}(2\lambda/\beta - 1)|_{\lambda=\beta} = P_k^{(0,2n)}(1) = 1. \tag{4.17}$$

Using (4.16) and (4.17) to make the appropriate substitutions in (4.13), we now obtain an upper and a lower bound for $\Gamma_{n,k}^{D'}$, which are expressible in terms of Jacobi polynomials, and hence are easily computable. These are given in Theorem 4.4 below.

Theorem 4.4. *$\Gamma_{n,k}^{D'}$ satisfies the inequalities*

$$\frac{\beta^n}{\{\sum_{j=0}^{k}(2n+2j+1)[P_j^{(0,2n)}(2/\beta-1)]^2\}^{\frac{1}{2}}} \leq \Gamma_{n,k}^{D'} \leq \frac{\beta^n}{P_k^{(0,2n)}(2/\beta-1)}. \tag{4.18}$$

By the assumption that $0 < \beta < 1$, we have $2/\beta - 1 > 1$. Thus, by Theorem A.1 in the appendix, the sequence $\{P_j^{(0,2n)}(2/\beta-1)\}_{j=0}^{\infty}$ is positive and monotonically increasing. We can use this to replace the lower bound on $\Gamma_{n,k}^{D'}$ by a weaker but more informative one. This is done in Corollary 4.5 below.

Corollary 4.5. *$\Gamma_{n,k}^{D'}$ satisfies the following weaker form of (4.18):*

$$\frac{1}{\sqrt{(k+1)(2n+2k+1)}}\frac{\beta^n}{P_k^{(0,2n)}(2/\beta-1)} \leq \Gamma_{n,k}^{D'} \leq \frac{\beta^n}{P_k^{(0,2n)}(2/\beta-1)}. \tag{4.19}$$

As can be seen from (4.19), the upper and lower bounds on $\Gamma_{n,k}^{D'}$ are very close to each other, and this implies that the upper bound it quite tight. This is so especially for moderate values of n and k, as will be demonstrated numerically later.

For the sake of completeness, we combine the sequence of results on the upper bounds in Theorem 4.6 below.

Theorem 4.6. *$\Gamma_{n,k}^{*}$, $\Gamma_{n,k}^{D}$, and $\Gamma_{n,k}^{D'}$ are related by the inequalities*

$$\Gamma_{n,k}^{*} \leq \Gamma_{n,k}^{D} \leq \Gamma_{n,k}^{D'} \leq \tilde{\Gamma}_{n,k}, \tag{4.20}$$

where

$$\tilde{\Gamma}_{n,k} \equiv \frac{\beta^n}{P_k^{(0,2n)}(2/\beta-1)} = \frac{\beta^{n+k}}{\sum_{j=0}^{k}\binom{k}{j}\binom{2n+k}{j}(1-\beta)^j}. \tag{4.21}$$

Finally, we note that the upper bound $\tilde{\Gamma}_{n,k}$ for $\Gamma^*_{n,k}$ is valid for $n = 0$ also when the matrix A in (1.1) has zero eigenvalues as well as positive ones. In this case the Jacobi polynomial $P_k^{(0,2n)}(z)$ reduces to the Legendre polynomial $P_k(z)$. This causes the upper bound to be slightly inferior to that obtained from the corresponding Chebyshev polynomial as in (4.3).

Note: Using Preposition 3 in [SafVa1], and the argument in the proof of Theorem 4.2, we can write, for $n = 1$ and arbitrary k and $D' = [0, \beta]$, $0 < \beta < 1$,

$$q^*(\lambda) = \frac{T_{k+1}((1-\eta)\lambda/\beta + \eta)}{T_{k+1}((1-\eta)/\beta + \eta)}$$

and

$$\Gamma_{1,k}^{D'} = \frac{1}{T_{k+1}((1-\eta)/\beta + \eta)},$$

where $\eta = -\cos(\pi/2(k+1))$.

4.2. *The Case $D = [-\beta, -\alpha]$, $0 < \alpha < \beta$*

Note that we have not demanded $\beta < 1$ for this case as 1 is not contained in D for any value of β. The treatment of this case is identical to that of the previous case. With the interval $[\alpha, \beta]$ in the previous case replaced by the interval $[-\beta, -\alpha]$ of the present case, Lemma 4.1 and Theorem 4.2 remain unchanged. Theorem 4.3 remains essentially the same except that $\phi_{n,k}(\beta)$ on the right hand side of (4.13) now reads $\phi_{n,k}(-\beta)$, and the ν_j have the same value as before. As for the polynomials orthogonal on $[-\beta, 0]$ with respect to the weight function λ^{2n}, they are now constant multiples of $P_j^{(0,2n)}(-2\lambda/\beta - 1)$. Denoting $[-\beta, 0]$ by D', (4.18) and (4.19) now read, respectively,

$$\frac{\beta^n}{\left\{\sum_{j=0}^{k}(2n+2j+1)\left[P_j^{(0,2n)}(-2/\beta - 1)\right]^2\right\}^{\frac{1}{2}}} \leq \Gamma_{n,k}^{D'} \leq \frac{\beta^n}{|P_k^{(0,2n)}(-2/\beta - 1)|} \tag{4.22}$$

and

$$\frac{1}{\sqrt{(k+1)(2n+2k+1)}} \frac{\beta^n}{|P_k^{(0,2n)}(-2/\beta - 1)|} \leq \Gamma_{n,k}^{D'} \leq \frac{\beta^n}{|P_k^{(0,2n)}(-2/\beta - 1)|}. \tag{4.23}$$

Similarly, (4.20) remains valid with (4.21) replaced by

$$\tilde{\Gamma}_{n,k} \equiv \frac{\beta^n}{|P_k^{(0,2n)}(-2/\beta - 1)|} = \frac{\beta^{n+k}}{\sum_{j=0}^{k} \binom{k}{j} \binom{2n+k}{j} (1+\beta)^j}. \tag{4.24}$$

As before, (4.23) follows from (4.22) by observing that the sequence $\{|P_j^{(0,2n)}(-2/\beta - 1)|\}_{j=0}^{\infty}$ is monotonically increasing by Theorem A.1 in the appendix, since $-2/\beta - 1 < -1$ for $\beta > 0$.

4.3. *The Case* $D = [\alpha, \beta]$, $\alpha < 0 < \beta < 1$

Since we are not able to make direct use of the theory of Haar subspaces in this case, we do not know whether Lemma 4.1 and Theorem 4.2 have analogs. We may, however, still use the orthogonal polynomials $\phi_{n,k}(\lambda)$ on $[\alpha, \beta]$ with respect to the weight function λ^{2n}. Consequently, (4.8) holds trivially. Theorem A.3 from the appendix this time applies separately in the two subintervals $[\alpha, 0]$ and $[0, \beta]$. As a result, (4.13) becomes

$$\{(\beta - \alpha) \sum_{j=0}^{k} |\phi_{n,j}(1)|^2/\nu_j\}^{-\frac{1}{2}} \leq \Gamma_{n,k}^D \leq \frac{\max\{|\alpha^n \phi_{n,k}(\alpha)|, |\beta^n \phi_{n,k}(\beta)|\}}{|\phi_{n,k}(1)|}, \tag{4.25}$$

with ν_j as given in (4.14).

Again, for arbitrary α and β the orthogonal polynomials $\phi_{n,k}(\lambda)$ are not known analytically, so that the bounds in (4.25) can be given numerically only. For the case $\alpha = -\beta$, however, the $\phi_{n,k}(\lambda)$ can be expressed in terms of Jacobi polynomials, see (A.5) in the appendix. We have,

$$\varphi_{n,k}(\lambda) = \begin{cases} P_\nu^{(0,n-1/2)}\,(2(\lambda/\beta)^2 - 1) & \text{if } k = 2\nu, \\ (\lambda/\beta) P_\nu^{(0,n+1/2)}\,(2(\lambda/\beta)^2 - 1) & \text{if } k = 2\nu + 1, \end{cases} \tag{4.26}$$

i.e., $\phi_{n,k}(\lambda)$ is an even or odd function of λ, depending on whether k is even or odd, respectively. As a result of (4.26), we obtain

$$\varphi_{n,k}(\beta) = 1 \quad \text{for all } n \text{ and } k, \tag{4.27}$$

$$\varphi_{n,k}(1) = \begin{cases} P_\nu^{(0,n-1/2)}\,(2/\beta^2 - 1) & \text{if } k = 2\nu, \\ \beta^{-1} P_\nu^{(0,n+1/2)}\,(2/\beta^2 - 1) & \text{if } k = 2\nu + 1, \end{cases} \tag{4.28}$$

and

$$\nu_j = \int_{-\beta}^{\beta} \lambda^{2n} |\varphi_{n,j}(\lambda)|^2 d\lambda = \frac{\beta^{2n+1}}{n + j + 1/2} \quad \text{for all } n \text{ and } j. \tag{4.29}$$

Combining all these in (4.25), we obtain

$$\frac{\beta^n}{\{\sum_{j=0}^{k}(2n+2j+1)|\phi_{n,j}(1)|^2\}^{1/2}} \le \Gamma_{n,k}^{D} \le \frac{\beta^n}{|\phi_{n,k}(1)|}. \tag{4.30}$$

Since $0 < \beta < 1$, $2/\beta^2 - 1 > 1$. Consequently Theorem A.2 applies, and we have that the sequence $\{\phi_{n,j}(1)\}_{j=0}^{\infty}$ is positive and monotonically increasing. With the help of this, we can replace (4.30) by the weaker but more informative form

$$\frac{1}{\sqrt{(k+1)(2n+2k+1)}}\frac{\beta^n}{|\phi_{n,k}(1)|} \le \Gamma_{n,k}^{D} \le \frac{\beta^n}{|\phi_{n,k}(1)|}. \tag{4.31}$$

In summary,

$$\Gamma_{n,k}^{*} \le \Gamma_{n,k}^{D} \le \tilde{\Gamma}_{n,k}, \tag{4.32}$$

where

$$\tilde{\Gamma}_{n,k} = \frac{\beta^n}{|\varphi_{n,k}(1)|} = \frac{\beta^{n+k}}{\sum_{j=0}^{\nu}\binom{\nu}{j}\binom{n+\mu-1/2}{j}(1-\beta^2)^j}, \tag{4.33}$$

with

$$\nu = \lfloor \frac{k}{2} \rfloor \quad \text{and} \quad \mu = \lfloor \frac{k+1}{2} \rfloor, \tag{4.34}$$

as follows from (4.28) and (A.7) and (A.8).

We note that the upper bound $\tilde{\Gamma}_{n,k}$ given by (4.33) and (4.34) can be improved somewhat as follows: By the fact that $|\lambda|^n$ is symmetric with respect to the origin in $D = [-\beta, \beta]$, we see that the solution $q^*(\lambda)$ of the min-max problem in (4.1) is even or odd depending on whether k is even or odd, respectively. Thus

$$\Gamma_{n,k}^{D} = \min_{q \in \mathcal{P}_k} \max_{|\lambda| \le \beta} |\lambda^n q(\lambda)| = \begin{cases} \min_{h \in P_\nu} \max_{0 \le \lambda \le \beta} |\lambda^n h(\lambda^2)| & \text{if } k = 2\nu, \\ \min_{h \in P_\nu} \max_{0 \le \lambda \le \beta} |\lambda^{n+1} h(\lambda^2)| & \text{if } k = 2\nu + 1. \end{cases} \tag{4.35}$$

Making now the change of variable $\lambda^2 = \tau$, we have

$$\Gamma_{n,k}^{D} = \begin{cases} \min_{h \in P_\nu} \max_{0 \le \tau \le \beta^2} |\tau^{n/2} h(\tau)| & \text{if } k = 2\nu, \\ \min_{h \in P_\nu} \max_{0 \le \tau \le \beta^2} |\tau^{(n+1)/2} h(\tau)| & \text{if } k = 2\nu + 1. \end{cases} \tag{4.36}$$

We finally employ Theorem 4.4 to obtain

$$\frac{\beta^n}{\{\sum_{j=0}^{\nu}(n+2j+1)[P_j^{(0,n)}(2/\beta^2-1)]^2\}^{\frac{1}{2}}} \le \Gamma_{n,2\nu}^{D} \le \frac{\beta^n}{P_\nu^{(0,n)}(2/\beta^2-1)}$$

$$\frac{\beta^{n+1}}{\{\sum_{j=0}^{\nu}(n+2j+2)[P_j^{(0,n+1)}(2/\beta^2-1)]^2\}^{\frac{1}{2}}} \le \Gamma_{n,2\nu+1}^{D} \le \frac{\beta^{n+1}}{P_\nu^{(0,n+1)}(2/\beta^2-1)}. \tag{4.37}$$

The upper bounds on $\Gamma_{n,k}^D$ can again be unified to read

$$\Gamma_{n,k}^D \le \tilde{\Gamma}_{n,k} = \frac{\beta^{n+k}}{\sum_{j=0}^{\nu} \binom{\nu}{j} \binom{n+\mu}{j} (1-\beta^2)^j} \tag{4.38}$$

with ν and μ as defined in (4.34). Comparing $\tilde{\Gamma}_{n,k}$ in (4.38) with $\tilde{\Gamma}_{n,k}$ in (4.33), we see that the former is slightly smaller than the latter.

In case $\alpha \neq -\beta$, but $\alpha > -1$, we can use the treatment involving the Jacobi polynomials to get an upper bound on $\Gamma_{n,k}^D$. For this we let $\hat{\beta} = \max(|\alpha|, \beta)$. Then $D = [\alpha, \beta] \subseteq [-\hat{\beta}, \hat{\beta}] = \hat{D}$. Thus

$$\Gamma_{n,k}^D \le \Gamma_{n,k}^{\hat{D}} = \min_{q \in \mathcal{P}_k} \max_{-\hat{\beta} \le \lambda \le \hat{\beta}} |\lambda^n q(\lambda)| \le \frac{\hat{\beta}^{n+k}}{\sum_{j=0}^{\nu} \binom{\nu}{j} \binom{n+\mu}{j} (1-\hat{\beta}^2)^j}, \tag{4.39}$$

as follows from (4.38). Here ν and μ are as in (4.34). Of course, this bound will be close to $\Gamma_{n,k}^D$ provided $|\alpha|$ and β are sufficiently close to each other.

4.4. *The Case* $D = \{\lambda : \lambda = i\xi, -\beta \le \xi \le \beta, \ \beta > 0 \text{ real}\}$

As before, we can immediately start using the orthogonal polynomials $\phi_{n,k}(\lambda)$ over D with respect to the weight function $|\lambda|^{2n}$. These polynomials are given by

$$\phi_{n,k}(\lambda) = \begin{cases} P_\nu^{(0,n-1/2)}(-2(\lambda/\beta)^2 - 1) & \text{if } k = 2\nu, \\ -i\,(\lambda/\beta) P_\nu^{(0,n+1/2)}(-2(\lambda/\beta)^2 - 1) & \text{if } k = 2\nu + 1. \end{cases} \tag{4.40}$$

With this we can easily verify that (4.27) holds, (4.29) holds in the sense that $\nu_j = \beta^{2n+1}/(n+j+1/2)$, and (4.28) holds with the argument $(2/\beta^2 - 1)$ replaced by $(-2/\beta^2 - 1)$. Consequently, (4.30) holds. In case $0 < \beta \le 1$, Theorem A.2 from the appendix applies, and we have that the sequence $\{|\phi_{n,j}(1)|\}_{j=0}^{\infty}$ is monotonically increasing. Thus (4.31) holds. Finally, (4.32) holds with

$$\tilde{\Gamma}_{n,k} = \frac{\beta^n}{|\varphi_{n,k}(1)|} = \frac{\beta^{n+k}}{\sum_{j=0}^{\nu} \binom{\nu}{j} \binom{n+\mu-1/2}{j} (1+\beta^2)^j}, \tag{4.41}$$

where ν and μ are as given in (4.34).

Going through the arguments in the paragraph following (4.34), we can improve the bound on $\Gamma_{n,k}^D$ in this case too. In fact, (4.37) holds with the argument $2/\beta^2 - 1$ replaced by $-2/\beta^2 - 1$. The

new upper bounds can again be unified to read

$$\tilde{\Gamma}_{n,k} = \frac{\beta^{n+k}}{\sum_{j=0}^{\nu} \binom{\nu}{j} \binom{n+\mu}{j} (1+\beta^2)^j}. \tag{4.42}$$

4.5. *The General Case*

Drawing on our experience with the previous cases, we now propose to use the appropriate orthogonal polynomials to construct upper and lower bounds for $\Gamma_{n,k}^D$ when D is arbitrary.

Let $\{\phi_{n,j}(\lambda)\}_{j=0}^{\infty}$ be the sequence of polynomials orthogonal with respect to the nonnegative weight function $|\lambda|^{2n}$, in the sense

$$\int_\Omega |\lambda|^{2n} \overline{\phi_{n,i}(\lambda)} \phi_{n,j}(\lambda) d\Omega = \nu_j \delta_{ij}. \tag{4.43}$$

Here Ω stands for D when D is a curve or a domain, or the boundary of D in case D is a domain. As a result, $d\Omega$ is the line element on Ω if Ω is a curve, or the area element if Ω is a domain. By (4.1) and Theorem A.4 in the appendix, we have

$$\{c \sum_{j=0}^{k} |\phi_{n,j}(1)|^2/\nu_j\}^{-\frac{1}{2}} \leq \Gamma_{n,k}^D \leq \frac{\max_{\lambda \in D} |\lambda^n \phi_{n,k}(\lambda)|}{|\phi_{n,k}(1)|}, \tag{4.44}$$

where

$$c = \int_\Omega d\Omega. \tag{4.45}$$

Of course, in order to determine these bounds we need to find the polynomials $\phi_{n,k}(\lambda)$ numerically, possibly through the 3-term recursion relation that they satisfy. In addition, this recursion relation needs to be determined numerically too.

Important simplifications take place when D is a line segment between α and β, where α and β can be complex, in general. Of course, the complex number 1 is assumed to be outside D. Making the change of variable $\lambda = \alpha + e^{i\theta}\xi$, where $\theta = \arg(\beta - \alpha)$, we realize that $\phi_{n,k}(\lambda)$ is actually a real polynomial in ξ orthogonal on the real interval $[0, (\beta - \alpha)e^{-i\theta}] \equiv I$ with respect to the weight function $|\lambda|^{2n} = |\alpha + e^{i\theta}\xi|^{2n}$. This weight function is either strictly monotonic on I or has only one relative minimum there. In both cases, we can invoke Theorem A.3 from the appendix to simplify $\max_{\lambda \in D} |\lambda^n \phi_{n,k}(\lambda)|$ that appears on the right hand side of (4.44). In case the weight function is

monotonic on I, let

$$\delta = \begin{cases} \alpha & \text{if } |\alpha| > |\beta|, \\ \beta & \text{if } |\beta| > |\alpha|. \end{cases} \tag{4.46}$$

Then

$$\max_{\lambda \in D} |\lambda^n \phi_{n,k}(\lambda)| = |\delta^n \phi_{n,k}(\delta)|. \tag{4.47}$$

In case the weight function is not monotonic on I,

$$\max_{\lambda \in D} |\lambda^n \phi_{n,k}(\lambda)| = \max\{|\alpha^n \phi_{n,k}(\alpha)|, |\beta^n \phi_{n,k}(\beta)|\}. \tag{4.48}$$

Further simplifications become possible when the origin is on the straight line containing the line segment D. First, assume that the origin is not in D, and let $|\alpha| \leq |\beta|$ without loss of generality. When $|\alpha| > 0$ we can replace $\phi_{n,k}(\lambda)$ by $P_k^{(0,2n)}(2\lambda/\beta - 1)$, the polynomial orthogonal on the line segment joining 0 and β with respect to the weight function $|\lambda|^{2n}$. Thus

$$\Gamma_{n,k}^D \leq \frac{|\beta|^n}{|P_k^{(0,2n)}(2/\beta - 1)|} = \frac{|\beta|^{n+k}}{|\sum_{j=0}^k \binom{k}{j} \binom{2n+k}{j} (1-\beta)^j|}. \tag{4.49}$$

Next, assume that $\alpha = -\beta$ so that the origin is at the center of D. For this case, we have

$$\Gamma_{n,k}^D \leq \frac{|\beta|^{n+k}}{|\sum_{j=0}^{\nu} \binom{\nu}{j} \binom{n+\mu}{j} (1-\beta^2)^j|}, \tag{4.50}$$

where ν and μ are as in (4.34). Lower bounds on $\Gamma_{n,k}^D$ can similarly be obtained with the help of the appropriate Jacobi polynomials. We should remember, though, that β in (4.46)-(4.50) is a complex number.

Finally, in cases where D is an ellipse with its semimajor axis along the real or the imaginary axis in the λ-plane, we can extend our previous results to obtain bounds on $\Gamma_{n,k}^*$ in conjunction with Bernstein's theorem, which is stated below.

Theorem 4.7. *Let $p(z)$ be a polynomial of degree at most k. Denote by $\mathcal{E}_\tau$ the ellipse with foci at ± 1, semimajor axis $\frac{1}{2}(\tau + \tau^{-1})$ and semiminor axis $\frac{1}{2}(\tau - \tau^{-1})$, where $\tau > 1$. Then*

$$\max_{z \in \mathcal{E}_\tau} |p(z)| \leq \tau^k \max_{x \in [-1,1]} |p(x)|. \tag{4.51}$$

As a result of this theorem, we see that if the foci of the ellipse are at α and β, $0 \leq \alpha < \beta < 1$ or at $-\beta$ and $-\alpha$, $0 \leq \alpha < \beta$, or at $\pm\beta$, $0 < \beta < 1$, or at $\pm i\beta$, $\beta > 0$, then the bounds $\tilde{\Gamma}_{n,k}$ given

in (4.21) or (4.24) or (4.38) or (4.42), respectively, need to be multiplied by τ^{n+k} for some $\tau > 1$, whose size depends on the size of ellipse. Of course, this will be so provided 1 lies outside the ellipse. The thinner the ellipse, the closer τ is to 1.

5 Appraisal of the Upper Bounds on $\Gamma^*_{n,k}$

From the definition of $\Gamma^*_{n,k}$ given in (3.13), it is clear that the sequence $\{\Gamma^*_{n,k}\}_{k=0}^{\infty}$ is monotonically decreasing for all spectra. It is also clear that, when $\rho(A) < 1$, the sequence $\{\Gamma^*_{n,k}\}_{n=0}^{\infty}$ is monotonically decreasing. We should, therefore, make sure that the upper bounds that we obtain for $\Gamma^*_{n,k}$ have these two characteristics. A cursory look at the expressions for $\bar{\Gamma}_{n,k}$ given in (4.21), (4.24), (4.38), and (4.42) for the different spectra reveals that both characteristics are possessed by the $\bar{\Gamma}_{n,k}$, in general. When $\rho(A) < 1$, the sequences $\{\bar{\Gamma}_{n,k}\}_{n=0}^{\infty}$ are monotonically decreasing for all cases considered. The sequences $\{\bar{\Gamma}_{n,k}\}_{k=0}^{\infty}$ are monotonically decreasing for the spectra contained in $D_1 = [\alpha, \beta]$, $0 \leq \alpha < \beta < 1$, $D_2 = [-\beta, -\alpha]$, $0 \leq \alpha < \beta$, $D_3 = [-\beta, \beta]$, $0 < \beta < 1$, and $D_4 = \{\lambda = i\xi : \ -\beta \leq \xi \leq \beta, \ \beta > 0 \text{ real }\}$, $\beta \leq 1$. Recall that $\bar{\Gamma}_{n,k}$ for D_4, which is given in (4.42), is valid for all β and not only for $\beta < 1$. For arbitrary β, the sequences $\{\bar{\Gamma}_{n,2\nu}\}_{\nu=0}^{\infty}$ and $\{\bar{\Gamma}_{n,2\nu+1}\}_{\nu=0}^{\infty}$ for D_4 are monotonically decreasing, as follows from (4.42) and Theorem A.2 in the appendix.

Let us first compare the $\bar{\Gamma}_{n,k}$ for the sets D_i, $i = 1, ..., 4$. It is seen by comparing (4.21) and (4.24) that, for a given spectral radius β, $\bar{\Gamma}_{n,k}$ for D_2 is smaller and decreases more quickly than that for D_1. Similarly, for a given spectral radius β, $\bar{\Gamma}_{n,k}$ for D_4 is smaller and decreases more quickly than that for D_3. For a given spectral radius, $\bar{\Gamma}_{n,k}$ is smallest and decreases most quickly for D_2.

We would now like to demonstrate by actual computation that the bounds $\bar{\Gamma}_{n,k}$ that were presented in Section 4 are very close to $\Gamma^D_{n,k}$. In all of our computations we picked $D_1 = [0, \beta]$, $D_2 = [-\beta, 0]$, $D_3 = [-\beta, \beta]$, and $D_4 = \{\lambda = i\xi : \ -\beta \leq \xi \leq \beta\}$, all with $\beta = 0.96$. In all cases we also computed the upper bounds obtained for (2.10) and (2.11) by Chebyshev polynomials, namely,

$$\Gamma^D_{n,k} \leq \frac{\beta^n}{T_k\left(\frac{2-\alpha-\beta}{\beta-\alpha}\right)} \equiv \Gamma^{Ch}_{n,k}, \tag{5.1}$$

for $D = [\alpha, \beta]$, $\alpha < \beta < 1$, thus covering D_1, D_2, and D_3, and

$$\Gamma^D_{n,k} \leq \frac{\beta^n}{|T_k\left(\frac{1}{\beta}\right)|} \equiv \Gamma^{Ch}_{n,k} \tag{5.2}$$

for $D = D_4$. (For D_1, D_2, and D_3 the inequality in (5.1) is actually an equality when $n = 0$ as follows from (4.3).) As mentioned previously, these bounds do not explain the behavior of $s_{n,k}$ for $n > 0$. They are given only for the sake of comparison. Finally, we computed the lower bounds on $\Gamma^D_{n,k}$ in order to verify that the upper bounds $\tilde{\Gamma}_{n,k}$ are indeed quite tight. All the computations reported in this section were done on an IBM-370 computer in double precision arithmetic.

Tables 5.1-5.4 contain the lower and upper bounds for $\Gamma^D_{n,k}$ and the Chebyshev polynomial bounds given in (5.1) and (5.2), for $n = 0, 50, 100$, and $k = 0, 2, 4, ..., 20$. Note the closeness of the lower and upper bounds which implies that both are close to $\Gamma^D_{n,k}$. Note also that both bounds decrease at an increasing rate as n increases.

6 Numerical Examples

In this section we give two numerical examples that provide ample support for the claims that were made in the previous sections. All the computations reported in this section were done on an IBM-370 computer by using the FORTRAN 77 code given in [Si3].

Example 6.1. Consider the linear system in (1.1), where the matrix A is the $N \times N$ tridiagonal matrix

$$A = \begin{bmatrix} \sigma & \tau & & & \\ \rho & \sigma & \tau & & \\ & \rho & \sigma & \tau & \\ & & \cdot & \cdot & \cdot \\ & & & \cdot & \cdot & \cdot \end{bmatrix}, \quad \rho, \sigma, \tau \text{ real.} \tag{6.1}$$

The eigenvalues of A are given by

$$\tilde{\lambda}_j = \sigma + 2\sqrt{\rho\tau}\cos\frac{j\pi}{N+1}, \quad j = 1, 2, ..., N.$$

It is seen that for large values of N there is a considerable amount of clustering of the eigenvalues near $\sigma + 2\sqrt{\rho\tau}$ and $\sigma - 2\sqrt{\rho\tau}$.

By adjusting the parameters ρ, σ, and τ we can cause the spectrum of A to be real and positive, or real and negative, or real and mixed, or pure imaginary, or complex in general. We can then test the upper bounds on $\Gamma^*_{n,k}$ given in Section 4. By Theorem 3.3, the easiest tests can be performed with RRE when the matrix A is normal, since for this case we have

$$W_{n,k} \equiv \frac{||r(s_{n,k})||}{||r(x_0)||} \leq \Gamma^*_{n,k} \leq \Gamma^D_{n,k} \leq \tilde{\Gamma}_{n,k}. \tag{6.2}$$

The $W_{n,k}$ for these tests were computed in extended double precision. The reason that we used such a high precision is that there is a considerable amount of loss of accuracy in the implementation of RRE (or any other vector extrapolation method) to obtain $s_{n,k}$ for large n and increasing k, especially when the spectrum of A is real and positive. All our numerical results clearly demonstrate that the upper bound $\tilde{\Gamma}_{n,k}$ for $\Gamma^D_{n,k}$ is actually very close to $W_{n,k}$, hence presents a true picture of the accuracy achieved in extrapolation with $s_{n,k}$, provided enough precision is used.

In our next experiment we compared cycling with $s_{n,k}$ for $n > 0$ to cycling with $s_{0,k}$, again using RRE. This time we did the computations in double precision only, causing round off to be considerable. Our numerical results for this example indicate that, for a prescribed level of accuracy, cycling with $s_{n,k}$, $n > 0$, can be much less costly than cycling with $s_{0,k}$, even in the presence of round off. The cost of cycling here is being measured in units of one iteration with (1.2). Since A is tridiagonal in this example, the cost of one such iteration is 3 vector additions and 3 scalar-vector multiplications, a total of 6 vector operations. As for the cost of computing $s_{n,k}$, it is made up of the cost of $n+k+1$ iterations and the overhead due to the implementation of RRE. This overhead is $\frac{1}{2}(k^2+5k+2)$ vector additions, $\frac{1}{2}(k^2+5k+1)$ scalar-vector multiplications, and $\frac{1}{2}(k^2+3k+2)$ scalar products. For linear systems, by taking advantage of the relation

$$u_m = A\,u_{m-1}, \quad m = 1, 2, \ldots, \tag{6.3}$$

we can reduce the cost by $2k$ vector additions. Thus, the overhead now becomes $\frac{1}{2}(k^2+k+1)$ vector additions, the number of scalar-vector multiplications and scalar products remaining as before. Since one scalar product is almost equivalent to one vector addition and one scalar-vector multiplication, we see that, rougly speaking, the overhead is k^2+2k+2 vector additions and k^2+4k+1 scalar-vector multiplications, a total of $2k^2+6k+3$ vector operations. Therefore, within each cycle, the computation of $s_{n,k}$ costs as much as $n+k+1+(2k^2+6k+3)/6$ iterations for the present example.

The computations were done for the following cases:

1. $\rho = \tau = \sigma/2$, $0 < \sigma < 1/2$. In this case $D_1 = [0, 2\sigma]$.
   Pick $\sigma = 0.48$, so that $D_1 = [0, 0.96]$.

2. $\rho = \tau = -\sigma/2$, $-\frac{1}{2} < \sigma < 0$. In this case $D_2 = [-2\sigma, 0]$.
   Pick $\sigma = -0.48$, so that $D_2 = [-0.96, 0]$.

3. $\rho = \tau < \frac{1}{2}$, $\sigma = 0$. In this case $D_3 = [-2\rho, 2\rho]$.
   Pick $\rho = 0.48$, so that $D_3 = [-0.96, 0.96]$.

4. $\rho = -\tau$, $0 < \rho < \frac{1}{2}$, $\sigma = 0$. In this case $D_4 = \{\lambda = i\xi : \ -2\rho \le \xi \le 2\rho\}$.

Pick $\rho = 0.48$, so that $D_4 = \{\lambda = i\xi : \ -0.96 \le \xi \le 0.96\}$.

Tables 6.1.1-6.1.4 provide the values of $W_{n,k}$ and $\tilde{\Gamma}_{n,k}$ for $k = 0, 1, ..., 20$ with $n = 80$. Figures 6.1.1-6.1.4 show $\log_{10}(\|r(x_i)\|/\|r(x_{init})\|)$, $0 \le i \le n$, and $\log_{10}(\|r(s_{n,k})\|/\|r(x_{init})\|)$, $1 \le k \le K$, for (i) $n = 0$ and $K = 10$ and (ii) $n = 50$ and $K = 10$, versus the cost of computing the x_i or the $s_{n,k}$ in the cycling mode. Here x_{init} is the initial vector given as $x_{init} = (1, 1/\sqrt{2}, 1/\sqrt{3}, ..., 1/\sqrt{N})^T$. The vector b in (1.1) and (1.2) is chosen to be zero so that the solution s is also zero.

Now in the cases treated in Figures 6.1.1-6.1.4 the matrix A is symmetric or antisymmetric. We next consider the case in which A is neither symmetric not antisymmetric. In the numerical experiment below we pick $\sigma = 0$, $\rho = 0.6$, and $\tau = 0.384$ so that the spectrum of A is contained in $[-0.96, 0.96]$. Since A is not normal, the norms of the vectors $u_n = A^n u_0$ do not behave like $(0.96)^n$ numerically. Their behavior is more like $(0.984)^n$, where $0.984 = \rho + \tau$. To see this we simply take $u_0 = (1, 1, ... 1)^T$ and actually compute $A^n u_0$. Table 6.1.5 provides the values of $W_{n,k}$ and $\tilde{\Gamma}_{n,k}$ for the interval $D' = [-0.984, 0.984]$, as if A were normal, for $k = 0, 1, ..., 20$, with $n = 80$. Although $\tilde{\Gamma}_{n,k}$ is only a heuristic estimate, it, nevertheless, is quite realistic. (The upper bound given in (3.16) becomes very pessimistic in this case as $\text{cond}_2(R) = (\rho/\tau)^{N-1}$ is of the order of 10^{193}. The effect of the dimension N on $\text{cond}_2(R)$ should be noted here. Even though ρ and τ may be nearly equal, a sufficiently large value of N can cause $\text{cond}_2(R)$ to be extremely large. This will also have an effect on the convergence behavior of $s_{n,k}$). Figure 6.1.5 shows $\log_{10}(\|r(x_i)\|/\|r(x_{init})\|)$, $0 \le i \le n$, and $\log_{10}(\|r(s_{n,k})\|/\|r(x_{init})\|)$, $1 \le k \le K$, for (i) $n = 0$ and $K = 10$ and (ii) $n = 50$ and $K = 10$, versus the cost of computing the x_i or the $s_{n,k}$ in the cycling mode, exactly as in Figures 6.1.1-6.1.4.

Example 6.2. Consider the 2-dimensional convection-diffusion equation

$$-\frac{\partial^2 u}{\partial x^2} - \frac{\partial^2 u}{\partial y^2} + \gamma\left(x\frac{\partial u}{\partial x} + y\frac{\partial u}{\partial y}\right) + \beta u = f \text{ in } \Omega,$$

$$u = g \text{ on } \partial\Omega,$$

where Ω is the unit square. This equation has been used as a test problem for vector extrapolation methods and Krylov subspace methods on nonsymmetric and/or indefinite systems. See, e.g., Gander, Golub, and Gruntz [GaGoGr].

Let $x_i = i\delta x$, $0 \le i \le M_x + 1$, and $y_j = j\delta y$, $0 \le j \le M_y + 1$, where $\delta x = 1/(M_x + 1)$ and $\delta y = 1/(M_y + 1)$ for some positive integers M_x and M_y. We discretize this equation by replacing all the partial derivations at (x_i, y_j) by central differences. If we now order the unknowns $u_{i,j}$, which are

the approximations to the corresponding $u(x_i, y_j)$, in the form $u_{11}, u_{12}, ..., u_{1M_y}, u_{21}, u_{22}, ..., u_{2M_y}, ...,$ $u_{M_x1}, u_{M_x2}, ..., u_{M_xM_y}$, then we obtain a linear system of equations with a block tridiagonal matrix. If $\beta = 0 = \gamma$, then we have the usual 5-point discretization scheme for Poisson's equation, in which case the matrix of the linear system is symmetric and positive definite. By increasing β in the negative direction we can make the matrix less and less positive definite and ultimately cause it to become indefinite. By picking $\gamma \neq 0$ we make the matrix nonsymmetric, the amount of asymmetry being directly related to the size of γ.

In our computations we picked $M_x = M_y = 30$ so that the number of unknowns is $N = M_xM_y = 900$. We also took $g = 0$ as our boundary condition and $f = 0$, causing the solution (both of the partial differential equation and of the difference equations) to be zero everywhere. For all our computations we took $x_{init} = (1, 1/\sqrt{2}, 1/\sqrt{3}, ..., 1/\sqrt{N})^T$ as our initial vector. The iterative technique used is the Jacobi method. The extrapolation method used in conjunction with this iterative technique is RRE.

If we write the linear system above as $Bx = d$, then B is of the form $B = \delta I - T$, where δ is a nonzero scalar and T has zero diagonal. This being the case, the matrix of iteration for the Jacobi method is simply $A = \delta^{-1}T = I - \delta^{-1}B$. By Theorem 2.4 in [Si2], the vector $s_{0,k}$ obtained by applying MPE (RRE) to the vector sequence $x_0, x_1, x_2, ...$, where $x_{j+1} = Ax_j + \delta^{-1}d$, is equivalent to that obtained by applying the k-step Arnoldi method (GCR = ORTHODIR = Axelson's method = GMRES) to the linear system $(I - A)x = \delta^{-1}d$, starting with the vector x_0 as the initial vector. But since $I - A = \delta^{-1}B$, this linear system is simply a constant multiple of $Bx = d$. As a result, $s_{0,k}$ with MPE (RRE) is, in fact, the vector obtained by applying the k-step Arnoldi method (GCR = ORTHODIR = Axelson's method = GMRES) to the linear system $Bx = d$, starting with x_0 as the initial vector.

In case the matrix B is of the form $B = D - T$, where D is a diagonal matrix that is invertible but $D \neq \delta I$ for any δ, and the matrix T has zero diagonal, the matrix of iteration for the Jacobi method is $A = D^{-1}T$. Again by Theorem 2.4 in [Si2] $s_{0,k}$ obtained by applying MPE (RRE) to the vector sequence $x_0, x_1, x_2, ...$, where $x_{j+1} = Ax_j + D^{-1}d$, is equivalent to that obtained by applying the k-step Arnoldi method (GCR = ORTHODIR = Axelson's method = GMRES) to the *diagonally preconditioned* linear system $D^{-1}Bx = D^{-1}d$, starting with the vector x_0.

In Figures 6.2.1-6.2.3 we show $\log_{10}||r(x_i)||$, $0 \leq i \leq n$, and $\log_{10}||r(s_{n,k})||$, $1 \leq k \leq K$, for (i) $n = 0$ and $K = 20$ and (ii) $n = 50$ and $K = 20$, versus the cost of computing the x_i or the $s_{n,k}$ in the cycling mode, exactly as before. The cost of one iteration this time is about 10 vector operations.

In Figure 6.2.1 we show the results obtained by picking $\gamma = 100$ and $\beta = 0$. The Jacobi method converges for this case. In about 500 iterations the residuals decrease by as much as 10^{-11}. Cycling with both $s_{0,20}$ and $s_{50,20}$ converges for this case, although the latter produces much better results for a prescribed amount of work.

In Figure 6.2.2 we show the results obtained by picking $\gamma = 100$ and $\beta = -100$. (In case $\gamma = 0$, such a large and negative value for β will cause the matrix B to become indefinite). The Jacobi method converges for this case too. In about 1000 iterations the residuals decrease by as much as 10^{-10}. Cycling with $s_{50,20}$ converges extremely quickly, while cycling with $s_{0,20}$ stalls after the first cycle.

In Figure 6.2.3, we show the results obtained by picking $\gamma = 125$ and $\beta = -100$. The Jacobi method now diverges. Cycling with $s_{50,20}$ converges extremely quickly, while cycling with $s_{0,20}$ stalls as in the previous case.

Observing that the matrix of the linear system above is consistently ordered, we can use the strategy that was proposed in [Si3, Section 7] to further reduce the computational cost, reducing the storage requirements by almost a half at the same time. According to this strategy, vector extrapolation methods are applied to the vector sequence obtained by using the double Jacobi iteration technique. With x_0 given, and A being the matrix of the Jacobi iteration method, the double Jacobi iteration technique produces the vectors $x_1, x_2, \ldots$, in accordance with

$$y = Ax_j + b$$
$$x_{j+1} = Ay + b$$
$$j = 0, 1, 2, \ldots .$$

We then expect cycling with the double Jacobi iteration using $s_{n,k}$ to produce results similar to those produced by cycling with the (single) Jacobi iteration using $s_{2n,2k}$. Obviously, the number of single Jacobi iterations actually performed in both cases is almost the same, although the computational cost and storage requirements for it are much lower with the double Jacobi iteration technique.

Figures 6.2.4-6.2.6 show $\log_{10}||r(x_i)||$, $0 \le i \le n$, $\log_{10}||r(s_{n,k})||$, $1 \le k \le K$, for (i) $n = 0$ and

$K = 10$ and (ii) $n = 25$ and $K = 10$, versus the cost of computing the x_i or the $s_{n,k}$ in the cycling mode, exactly as before.

Figures 6.2.4-6.2.6 show the results obtained by picking $\gamma = 100$ and $\beta = 0$, $\gamma = 100$ and $\beta = -100$, and $\gamma = 125$ and $\beta = -100$, respectively, as before, the conclusions being also as before.

Note that the cost of a double Jacobi iteration is twice that of a single Jacobi iteration, namely, 20 vector operations. Therefore, when comparing the costs in Figures 6.2.4-6.2.6 with the corresponding costs in Figures 6.2.1-6.2.3, the former should be doubled.

Finally, we mention that for both Example 6.1 and Example 6.2, MPE can be used instead of RRE, the results obtained being very similar.

Appendix

1. *A Collection of Useful Formulas and Results for Jacobi Polynomials*

The Jacobi polynomials $P_k^{(\alpha,\beta)}(x)$ are defined by

$$P_k^{(\alpha,\beta)}(x) = \sum_{j=0}^{k} \binom{k+\alpha}{k-j} \binom{k+\beta}{j} \left(\frac{x-1}{2}\right)^j \left(\frac{x+1}{2}\right)^{k-j} \tag{A.1}$$

with $\alpha > -1$ and $\beta > -1$. They are orthogonal with respect to the weight function $w(x) = (1-x)^\alpha(1+x)^\beta$ on $[-1,1]$, i.e.,

$$\int_{-1}^{1} (1-x)^\alpha (1+x)^\beta P_m^{(\alpha,\beta)}(x) P_k^{(\alpha,\beta)}(x) dx =$$

$$\delta_{mk} \frac{2^{\alpha+\beta+1}}{2k+\alpha+\beta+1} \frac{\Gamma(k+\alpha+1)\Gamma(k+\beta+1)}{\Gamma(k+1)\Gamma(k+\alpha+\beta+1)}, \tag{A.2}$$

where δ_{mk} is the Kronecker delta. $P_k^{(\alpha,\beta)}(x)$ are normalized such that

$$P_k^{(\alpha,\beta)}(1) = \binom{k+\alpha}{k}. \tag{A.3}$$

Polynomials orthogonal on $[a,b]$ with respect to the weight function $w(x) = (b-x)^\alpha(x-a)^\beta$ are

$$p_k(x) = P_k^{(\alpha,\beta)}\left(2\frac{x-a}{b-a} - 1\right). \tag{A.4}$$

Polynomials orthogonal on $[-1,1]$ with respect to the weight function $w(x) = |x|^{2n}$ are given by

$$p_k(x) = \begin{cases} P_\nu^{(0,n-1/2)}(2x^2-1) & \text{if } k = 2\nu \\ x P_\nu^{(0,n+1/2)}(2x^2-1) & \text{if } k = 2\nu+1. \end{cases} \tag{A.5}$$

The normalization condition given in (A.3) is the one that has been widely accepted in the literature of orthogonal polynomials. Thus (A.1) - (A.3) can be found in many books. See e.g., [AbSteg, Chapter 22] or [Sz]. For (A.5) see [Sz, pp. 59-60].

Theorem A.1. *For $x > 1$ or $x < -1$, with x fixed otherwise, the sequence $\{|P_k^{(\alpha,\beta)}(x)|\}_{k=0}^{\infty}$ is monotonically increasing.*

Proof. We start with the case $x > 1$. First, all the terms in the summation on the right hand side of (A.1) are positive for $x > 1$. Next, the jth term of $P_k^{(\alpha,\beta)}(x)$ in (A.1) is strictly less than the

corresponding term of $P_{k+1}^{(\alpha,\beta)}(x)$. The result now follows. As for $x < -1$, we first recall that

$$P_k^{(\alpha,\beta)}(-x) = (-1)^k P_k^{(\beta,\alpha)}(x), \tag{A.6}$$

and then apply the result for $x > 1$, which we have already proved, to the polynomials $P_k^{(\beta,\alpha)}(-x)$. □

Theorem A.2. *The polynomials $p_k(x)$ that are defined in (A.5) are such that, for n real and $|x| > 1$, or for x pure imaginary and $|x| \geq 1$, the sequence $\{|p_k(x)|\}_{k=0}^{\infty}$ is monotonically increasing. For x pure imaginary and $|x| < 1$ the sequences $\{|p_{2\nu}(x)|\}_{\nu=0}^{\infty}$ and $\{|p_{2\nu+1}(x)|\}_{\nu=0}^{\infty}$ are monotonically increasing.*

Proof. We observe that, by proper manipulation of (A.1), $p_k(x)$ can be expressed in the unified form

$$p_k(x) = \sum_{j=0}^{\nu} \binom{\nu}{j} \binom{n+\mu-1/2}{j} (x^2-1)^j x^{k-2j}, \tag{A.7}$$

where

$$\nu = \lfloor \frac{k}{2} \rfloor \text{ and } \mu = \lfloor \frac{k+1}{2} \rfloor. \tag{A.8}$$

Note that both ν and μ are monotonically nondecreasing in k, and that one of them is always increasing. Letting now x be real and $x > 1$, we see that all the terms in the summation on the right hand side of (A.7) are positive. Next, the jth term of $p_k(x)$ in (A.7) is strictly less than the corresponding term of $p_{k+1}(x)$. The result now follows for $x > 1$. For $x < -1$, we note that $p_k(-x) = (-1)^k p_k(x)$, and apply the result for $x > 1$, which we have already proved, to the polynomials $p_k(-x)$. For the case in which x is pure imaginary, i.e., $x = i\xi$, ξ real, the factor $(x^2-1)^j x^{k-2j}$ in the jth term of $p_k(x)$ becomes $i^k(\xi^2+1)^j \xi^{k-2j}$. The proof for the case $|x| \geq 1$ can now be completed as before. The proof of the case $|x| < 1$ can be done by employing Theorem A.1 in conjunction with (A.5). □

2. *A Result on Monotonic Weight Functions*

Theorem A.3. *Let $\{p_n(x)\}_{n=0}^{\infty}$ be the sequence of polynomials orthogonal on $[a,b]$ for finite b with respect to the nonnegative weight function $w(x)$. Assume that $w(x)$ is nondecreasing on $[a,b]$. Then the functions $\sqrt{w(x)}|p_n(x)|$ attain their maximum on $[a,b]$ for $x = b$. A corresponding statement holds for any subinterval $[x_0,b]$ of $[a,b]$ where $w(x)$ is nondecreasing.*

This theorem is stated and proved in [Sz, p. 163, Theorem 7.2].

3. *A Lower Bound for a Best Polynomial l_∞- Approximation Problem*

Theorem A.4. *Let $\{p_n(z)\}_{n=0}^{\infty}$ be the sequence of orthogonal polynomials on a compact set Ω of the complex z-plane with respect to the real nonnegative weight function $w(z)$ on Ω, i.e.,*

$$\int_\Omega w(z)\overline{p_m(z)}p_n(z)d\Omega = \delta_{m,n}, \tag{A.9}$$

where $d\Omega$ stands for the area element if Ω is a domain D, and for the line element if Ω is the boundary of a domain D or an arbitrary rectifiable curve. Let $\phi^*(z)$ be the solution of the constrained min-max problem

$$\min_{\phi} \max_{z\in\Omega} |\sqrt{w(z)}\phi(z)|, \ \phi(z) \text{ polynomial of degree } \leq k,$$

$$\text{subject to } M(\phi) = 1, \tag{A.10}$$

where M is a bounded linear functional on the space of functions continous on Ω. Then

$$\max_{z\in\Omega} |\sqrt{w(z)}\phi^*(z)| \ \geq \{c\sum_{j=0}^{k} |M(p_j)|^2\}^{-1/2}, \ c = \int_\Omega d\Omega. \tag{A.11}$$

Proof. We start by observing that, for any function $f(z)$ that is continuous on Ω, we have

$$\max_{z\in\Omega} |f(z)| \geq \{c^{-1}\int_\Omega |f(z)|^2 d\Omega\}^{1/2}. \tag{A.12}$$

Letting now $f(z) = \sqrt{w(z)}\phi(z)$ in (A.12), where $\phi(z)$ is a polynomial of degree at most k satisfying $M(\phi) = 1$, and minimizing both sides of (A.12) with respect to ϕ, we obtain

$$\max_{z\in\Omega} |\sqrt{w(z)}\phi^*(z)| \geq \min_{M(\phi)=1} \{c^{-1}\int_\Omega w(z)|\phi(z)|^2 d\Omega\}^{1/2}. \tag{A.13}$$

Since $\phi(z)$ is a polynomial of degree k, it can be written as

$$\phi(z) = \sum_{i=0}^{k} \alpha_i p_i(z), \tag{A.14}$$

so that the minimization problem on the right hand side of (A.13) becomes

$$\underset{\alpha_i}{\text{minimize}} \sum_{i=0}^{k} |\alpha_i|^2$$

$$\text{subject to } \sum_{i=0}^{k} \alpha_i M(p_i) = 1. \tag{A.15}$$

The solution of (A.15) can be achieved, e.g., by using the method of Lagrange multipliers, and is given by

$$\alpha_j = \frac{\overline{M(p_j)}}{\sum_{i=0}^{k} |M(p_i)|^2}, \; j = 0, 1, ..., k. \tag{A.16}$$

Combining (A.16) with (A.13) - (A.15), (A.11) follows. □

Obviously, in case $\Omega = [a, b]$, a finite real interval, we have $d\Omega = dx$ and $c = b - a$.

Also, if M is a point evaluation functional, i.e., $M(\phi) = \phi(\xi)$ for some ξ, then $M(p_i) = p_i(\xi)$ in (A.11).

Acknowledgements

The authors would like to thank Prof. Moshe Israeli for useful conversations and Prof. Allan Pinkus for drawing their attention to some of the references cited in this work. They also thank Mark L. Celestina for producing the graphs given in Section 6.

References

[AbSteg] M. Abramowitz and I.A. Stegun, *Handbook of Mathematical Functions*, National Bureau of Standards, Applied Mathematics Series, No. 55, Government Printing Office, Washington, D.C., 1964.

[Ar] W.E. Arnoldi, The principle of minimized iterations in the solution of the matrix eigenvalue problem, *Quart. Appl. Math.* **9** (1951), pp. 17-29.

[Ax] O. Axelsson, Conjugate gradient type methods for unsymmetric and inconsistent systems of linear equations, *Lin. Alg. Appl.* **29** (1980), pp. 1-16.

[CaJa] S. Cabay and L.W. Jackson, A polynomial extrapolation method for finding limits and antilimits of vector sequences, *SIAM J. Numer. Anal.* **13** (1976), pp.734-752.

[Ch] E.W. Cheney, *Introduction to Approximation Theory*, Second Edition, Chelsea, 1982.

[CoGo] P. Concus and G.H. Golub, A generalized conjugate gradient method for nonsymmetric systems of linear equations, in: R. Glowinski and J.L. Lions, Eds., *Proc. Second International Symposium on Computing Methods in Applied Sciences and Engineering*, IRIA, Paris, Dec. 1975. Lecture Notes in Economics and Mathematical Systems 134 (Springer, Berlin, 1976), pp. 56-65.

[Ed] R.P. Eddy, Extrapolating to the limit of a vector sequence, in: P.C.C. Wang, Ed., *Information Linkage between Applied Mathematics and Industry*, (Academic Press, New York, 1979), pp. 387-396.

[EiElSc] S.C. Eisenstat, H.C. Elman, and M.H. Schultz, Variational iterative methods for nonsymmetric systems of linear equations, *SIAM J. Numer. Anal.* **20** (1983), pp. 345-357.

[FoSi] W.F. Ford and A. Sidi, Recursive algorithms for vector extrapolation methods, *Appl. Numer. Math.* **4** (1988), pp. 477-489.

[FrR] R. Freund and S. Ruscheweyh, On a class of Chebyshev approximation problems which arise in connection with a conjugate gradient type method, *Numer. Math.* **48** (1986), pp. 525-542.

[GaGoGr] W. Gander, G.H. Golub, and D. Gruntz, Solving linear equations by extrapolation, Manuscript NA-89-11, Stanford University, Stanford, California (October 1989).

[HSti] M. Hestenes and E. Stiefel, Methods of conjugate gradients for solving linear systems, *J. Res. N.B.S.* **49** (1952), pp. 409-436.

[KStei] S. Kaniel and J. Stein, Least-square acceleration of iterative methods for linear equations, *J. Optim. Theory Appl.* **14** (1974), pp. 431-437.

[L] G.G. Lorentz, Approximation by incomplete polynomials (problems and results), in: E.B. Saff and R.S. Varga, eds., *Padé and Rational Approximations: Theory and Applications*, (Academic Press, New York, 1977), pp. 289-302.

[M] M. Mešina, Convergence acceleration for the iterative solution of the equations $X = AX + f$, *Comp. Meth. Appl. Mech. Eng.* **10** (1977), pp. 165-173.

[Saa] Y. Saad, Krylov subspace methods for solving large unsymmetric linear systems, *Math. Comp.* **37** (1981), pp. 105-126.

[SaaSc] Y. Saad and M.H. Schultz, A generalized minimal residual algorithm for solving nonsymmetric linear systems, *SIAM J. Sci. Stat. Computing* **7** (1986), pp. 856-869.

[SafVa1] E.B. Saff and R.S. Varga, On incomplete polynomials, in: L. Collatz, G. Meinardus, and H. Werner, eds., *Numerische Methoden der Approximationstheorie*, Band 4, ISNM, Vol.42, (Birkhäuser, Basel, 1978), pp. 281-298.

[SafVa2] E.B. Saff and R.S. Varga, Uniform approximation by incomplete polynomials, *Internat. J. Math. and Math. Sci.* **1** (1978), pp. 407-420.

[SafVa3] E.B. Saff and R.S. Varga, The sharpness of Lorentz's theorem on incomplete polynomials, *Trans. Amer. Math. Soc.*, **249** (1979), pp. 163-186.

[SafVa4] E.B. Saff and R.S. Varga, Incomplete polynomials II, *Pacific J. Math.* **92** (1981), pp. 161-172.

[Si1] A. Sidi, Convergence and stability properties of minimal polynomial and reduced rank extrapolation algorithms, *SIAM J. Numer. Anal.* **23** (1986), pp. 197-209.

[Si2] A. Sidi, Extrapolation vs. projection methods for linear systems of equations, *J. Comp. Appl. Math.* **22** (1988), pp. 71-88.

[Si3] A. Sidi, Efficient implementation of minimal polynomial and reduced rank extrapolation methods, *J. Comp. Appl. Math.* **36** (1991), pp. 305-337.

[SiB] A. Sidi and J. Bridger, Convergence and stability analyses for some vector extrapolation methods in the presence of defective iteration matrices, *J. Comp. Appl. Math.* **22** (1988), pp. 35-61.

[SiFoSm] A. Sidi, W.F. Ford, and D.A.Smith, Acceleration of convergence of vector sequences, *SIAM J. Numer. Anal.* **23** (1986), pp. 178-196.

[SmFoSi] D.A. Smith, W.F. Ford, and A. Sidi, Extrapolation methods for vector sequences, *SIAM Rev.* **29** (1987), pp. 199-233.

[Sti] E.L. Stiefel, Relaxationsmethoden bester Strategie zur losung linearer Gleichungssystems, *Comment. Math. Helv.* **29** (1955), pp. 157-179.

[Sz] G. Szegö, *Orthogonal Polynomials*, American Mathematical Society Colloquium Publications, Vol. 23, Providence, Rhode Island, 1939.

[Va] R.S. Varga, *Matrix Iterative Analysis*, Prentice-Hall, Englewood Cliffs, NJ, 1962.

[Vi] P.K.W. Vinsome, Orthomin, an iterative method for solving sparse sets of simultaneous linear equations, in: *Proc. Fourth Symposium on Reservoir Simulation*, Society of Petroleum Engineers of AIME, 1976, pp. 149-159.

[W] O. Widlund, A Lanczos method for a class of nonsymmetric systems of linear equations, *SIAM J. Numer. Anal.* **15** (1978), pp. 801-812.

[YJe] D.M. Young and K.C. Jea, Generalized conjugate gradient acceleration of nonsymmetrizable iterative methods, *Lin. Alg. Appl.* **34** (1980), pp. 159-194.

k	$n=0$			$n=50$			$n=100$		
	$\bar{\Gamma}'_{n,k}$	$\bar{\Gamma}_{n,k}$	$\bar{\Gamma}^{Cb}_{n,k}$	$\bar{\Gamma}'_{n,k}$	$\bar{\Gamma}_{n,k}$	$\Gamma^{Cb}_{n,k}$	$\bar{\Gamma}'_{n,k}$	$\bar{\Gamma}_{n,k}$	$\Gamma^{Cb}_{n,k}$
0	1.00D+00	1.00D+00	1.00D+00	1.29D-02	1.30D-01	1.30D-01	1.19D-03	1.69D-02	1.69D-02
2	2.83D-01	7.93D-01	7.42D-01	6.46D-04	6.88D-03	9.64D-02	2.15D-05	3.13D-04	1.25D-02
4	1.24D-01	5.00D-01	3.80D-01	7.71D-05	8.66D-04	4.94D-02	1.24D-06	1.86D-05	6.41D-03
6	5.66D-02	2.77D-01	1.74D-01	1.30D-05	1.52D-04	2.26D-02	1.16D-07	1.78D-06	2.94D-03
8	2.57D-02	1.43D-01	7.79D-02	2.66D-06	3.25D-05	1.01D-02	1.42D-08	2.25D-07	1.31D-03
10	1.16D-02	7.16D-02	3.47D-02	6.22D-07	7.85D-06	4.50D-03	2.12D-09	3.42D-08	5.85D-04
12	5.20D-03	3.50D-02	1.54D-02	1.60D-07	2.08D-06	2.00D-03	3.61D-10	5.97D-09	2.60D-04
14	2.32D-03	1.68D-02	6.85D-03	4.39D-08	5.88D-07	8.90D-04	6.85D-11	1.16D-09	1.16D-04
16	1.04D-03	8.00D-03	3.04D-03	1.28D-08	1.75D-07	3.95D-04	1.42D-11	2.43D-10	5.14D-05
18	4.62D-04	3.77D-03	1.35D-03	3.89D-09	5.47D-08	1.76D-04	3.13D-12	5.48D-11	2.28D-05
20	2.06D-04	1.77D-03	6.01D-04	1.23D-09	1.77D-08	7.81D-05	7.35D-13	1.31D-11	1.01D-05

Table 5.1. Bounds for $\Gamma^D_{n,k}$ when $D = [0, \beta]$ with $\beta = 0.96$.

$\bar{\Gamma}'_{n,k}$: the lower bound defined in (4.18),

$\bar{\Gamma}_{n,k}$: the upper bound defined in (4.18),

$\Gamma^{Cb}_{n,k}$: the Chebyshev bound defined in (5.1).

Note that $\Gamma^D_{0,k} = \Gamma^{Cb}_{0,k}$ for this case.

k	$n=0$			$n=50$			$n=100$		
	$\tilde{\Gamma}'_{n,k}$	$\tilde{\Gamma}_{n,k}$	$\tilde{\Gamma}^{Ch}_{n,k}$	$\tilde{\Gamma}'_{n,k}$	$\tilde{\Gamma}_{n,k}$	$\Gamma^{Ch}_{n,k}$	$\tilde{\Gamma}'_{n,k}$	$\tilde{\Gamma}_{n,k}$	$\Gamma^{Ch}_{n,k}$
0	1.00D+00	1.00D+00	1.00D+00	1.29D-02	1.30D-01	1.30D-01	1.19D-03	1.69D-02	1.69D-02
2	3.20D-02	7.27D-02	5.55D-02	5.79D-07	5.93D-06	7.21D-03	1.38D-08	1.97D-07	9.37D-04
4	9.12D-04	2.78D-03	1.54D-03	1.44D-10	1.50D-09	2.00D-04	9.21D-13	1.33D-11	2.60D-05
6	2.56D-05	9.36D-05	4.29D-05	8.28D-14	8.81D-13	5.57D-06	1.48D-16	2.16D-15	7.23D-07
8	7.14D-07	2.99D-06	1.19D-06	8.27D-17	8.95D-16	1.55D-07	4.26D-20	6.28D-19	2.01D-08
10	1.99D-08	9.25D-08	3.31D-08	1.23D-19	1.36D-18	4.30D-09	1.90D-23	2.83D-22	5.58D-10
12	5.54D-10	2.81D-09	9.19D-10	2.52D-22	2.82D-21	1.19D-10	1.20D-26	1.80D-25	1.55D-11
14	1.54D-11	8.41D-11	2.55D-11	6.62D-25	7.53D-24	3.31D-12	1.00D-29	1.52D-28	4.31D-13
16	4.28D-13	2.50D-12	7.09D-13	2.14D-27	2.48D-26	9.21D-14	1.07D-32	1.63D-31	1.20D-14
18	1.19D-14	7.35D-14	1.97D-14	8.28D-30	9.72D-29	2.56D-15	1.40D-35	2.15D-34	3.32D-16
20	3.31D-16	2.15D-15	5.47D-16	3.73D-32	4.44D-31	7.11D-17	2.19D-38	3.40D-37	9.23D-18

Table 5.2. Bounds for $\Gamma^D_{n,k}$ when $D = [-\beta, 0]$ with $\beta = 0.96$.

$\tilde{\Gamma}'_{n,k}$: the lower bound defined in (4.22),

$\tilde{\Gamma}_{n,k}$: the upper bound defined in (4.22),

$\Gamma^{Ch}_{n,k}$: the Chebyshev bound defined in (5.1).

Note that $\Gamma^D_{0,k} = \Gamma^{Ch}_{0,k}$ for this case.

k	$n=0$			$n=50$			$n=100$		
	$\tilde{\Gamma}'_{n,k}$	$\tilde{\Gamma}_{n,k}$	$\tilde{\Gamma}^{Ch}_{n,k}$	$\tilde{\Gamma}'_{n,k}$	$\tilde{\Gamma}_{n,k}$	$\Gamma^{Ch}_{n,k}$	$\tilde{\Gamma}'_{n,k}$	$\tilde{\Gamma}_{n,k}$	$\Gamma^{Ch}_{n,k}$
0	1.00D+00	1.00D+00	1.00D+00	1.82D-02	1.30D-01	1.30D-01	1.68D-03	1.69D-02	1.69D-02
2	4.42D-01	8.55D-01	8.55D-01	3.24D-03	2.39D-02	1.11D-01	1.71D-04	1.74D-03	1.44D-02
4	2.41D-01	6.44D-01	5.75D-01	8.31D-04	6.38D-03	7.47D-02	2.83D-05	2.94D-04	9.70D-03
6	1.38D-01	4.44D-01	3.45D-01	2.55D-04	2.03D-03	4.48D-02	6.02D-06	6.37D-05	5.82D-03
8	7.92D-02	2.90D-01	1.98D-01	8.78D-05	7.18D-04	2.57D-02	1.49D-06	1.61D-05	3.34D-03
10	4.54D-02	1.83D-01	1.12D-01	3.26D-05	2.74D-04	1.46D-02	4.14D-07	4.54D-06	1.89D-03
12	2.58D-02	1.13D-01	6.33D-02	1.28D-05	1.11D-04	8.22D-03	1.25D-07	1.39D-06	1.07D-03
14	1.47D-02	6.89D-02	3.56D-02	5.26D-06	4.65D-05	4.63D-03	4.01D-08	4.54D-07	6.01D-04
16	8.30D-03	4.15D-02	2.00D-02	2.23D-06	2.02D-05	2.60D-03	1.36D-08	1.56D-07	3.38D-04
18	4.69D-03	2.48D-02	1.13D-02	9.77D-07	9.02D-06	1.46D-03	4.81D-09	5.61D-08	1.90D-04
20	2.65D-03	1.47D-02	6.34D-03	4.37D-07	4.12D-06	8.24D-04	1.76D-09	2.09D-08	1.07D-04

Table 5.3. Bounds for $\Gamma^D_{n,k}$ when $D = [-\beta, \beta]$ with $\beta = 0.96$.

$\tilde{\Gamma}'_{n,k}$: the lower bound defined in (4.37),

$\tilde{\Gamma}_{n,k}$: the upper bound defined in (4.37),

$\Gamma^{Ch}_{n,k}$: the Chebyshev bound defined in (5.1).

Note that $\Gamma^D_{0,k} = \Gamma^{Ch}_{0,k}$ for this case.

k	$n=0$			$n=50$			$n=100$		
	$\tilde{\Gamma}'_{n,k}$	$\tilde{\Gamma}_{n,k}$	$\tilde{\Gamma}^{Ch}_{n,k}$	$\tilde{\Gamma}'_{n,k}$	$\tilde{\Gamma}_{n,k}$	$\Gamma^{Ch}_{n,k}$	$\tilde{\Gamma}'_{n,k}$	$\tilde{\Gamma}_{n,k}$	$\Gamma^{Ch}_{n,k}$
0	1.00D+00	1.00D+00	1.00D+00	1.82D-02	1.30D-01	1.30D-01	1.68D-03	1.69D-02	1.69D-02
2	1.79D-01	3.15D-01	3.15D-01	1.66D-04	1.21D-03	4.10D-02	7.85D-06	7.97D-05	5.32D-03
4	3.02D-02	6.86D-02	5.24D-02	2.92D-06	2.16D-05	6.80D-03	7.20D-08	7.38D-07	8.83D-04
6	4.98D-03	1.34D-02	8.48D-03	7.41D-08	5.59D-07	1.10D-03	9.72D-10	1.01D-08	1.43D-04
8	8.13D-04	2.47D-03	1.37D-03	2.42D-09	1.86D-08	1.78D-04	1.71D-11	1.79D-10	2.32D-05
10	1.32D-04	4.45D-04	2.22D-04	9.50D-11	7.42D-10	2.89D-05	3.71D-13	3.91D-12	3.75D-06
12	2.15D-05	7.85D-05	3.60D-05	4.33D-12	3.44D-11	4.67D-06	9.45D-15	1.00D-13	6.07D-07
14	3.49D-06	1.37D-05	5.82D-06	2.22D-13	1.79D-12	7.56D-07	2.76D-16	2.96D-15	9.82D-08
16	5.66D-07	2.36D-06	9.42D-07	1.26D-14	1.03D-13	1.22D-07	9.02D-18	9.77D-17	1.59D-08
18	9.17D-08	4.05D-07	1.52D-07	7.80D-16	6.49D-15	1.98D-08	3.26D-19	3.56D-18	2.57D-09
20	1.49D-08	6.90D-08	2.47D-08	5.19D-17	4.38D-16	3.20D-09	1.29D-20	1.42D-19	4.16D-10

Table 5.4. Bounds for $\Gamma^D_{n,k}$ when $D = \{\lambda = i\xi : \beta \le \xi \le \beta\}$ with $\beta = 0.96$.

$\tilde{\Gamma}'_{n,k}$: the lower bound defined in (4.37) with $2/\beta^2 - 1$ replaced by $-2/\beta^2 - 1$,

$\tilde{\Gamma}_{n,k}$: the upper bound defined in (4.42),

$\Gamma^{Ch}_{n,k}$: the Chebyshev bound defined in (5.2).

k	$W_{80,k}$	$\tilde{\Gamma}_{80,k}$	k	$W_{80,k}$	$\tilde{\Gamma}_{80,k}$	k	$W_{80,k}$	$\tilde{\Gamma}_{80,k}$	k	$W_{80,k}$	$\tilde{\Gamma}_{80,k}$
0	9.70D-03	3.82D-02	0	3.32D-04	3.82D-02	0	5.15D-03	3.82D-02	0	3.23D-04	3.82D-02
1	5.31D-04	4.92D-03	1	1.22D-06	1.16D-04	1	4.94D-03	3.66D-02	1	2.22D-04	3.66D-02
2	9.62D-05	1.01D-03	2	8.11D-09	6.93D-07	2	3.15D-04	4.79D-03	2	2.30D-06	2.25D-04
3	2.35D-05	2.61D-04	3	7.72D-11	6.15D-09	3	2.72D-04	4.55D-03	3	1.58D-06	2.13D-04
4	6.79D-06	7.82D-05	4	9.49D-13	7.19D-11	4	5.82D-05	9.46D-04	4	2.93D-08	2.58D-06
5	2.19D-06	2.59D-05	5	1.43D-14	1.04D-12	5	4.75D-05	8.91D-04	5	1.99D-08	2.42D-06
6	7.70D-07	9.29D-06	6	2.52D-16	1.78D-14	6	1.41D-05	2.34D-04	6	5.26D-10	4.33D-08
7	2.88D-07	3.54D-06	7	5.11D-18	3.50D-16	7	1.11D-05	2.19D-04	7	3.56D-10	4.01D-08
8	1.13D-07	1.42D-06	8	1.16D-19	7.80D-18	8	3.94D-06	6.68D-05	8	1.21D-11	9.46D-10
9	4.66D-08	5.90D-07	9	2.94D-21	1.93D-19	9	3.04D-06	6.20D-05	9	8.13D-12	8.67D-10
10	1.99D-08	2.55D-07	10	8.13D-23	5.26D-21	10	1.22D-06	2.10D-05	10	3.35D-13	2.53D-11
11	8.72D-09	1.13D-07	11	2.44D-24	1.55D-22	11	9.28D-07	1.94D-05	11	2.24D-13	2.29D-11
12	3.93D-09	5.17D-08	12	7.89D-26	4.96D-24	12	4.07D-07	7.09D-06	12	1.08D-14	7.90D-13
13	1.82D-09	2.41D-08	13	2.73D-27	1.69D-25	13	3.07D-07	6.52D-06	13	7.19D-15	7.08D-13
14	8.56D-10	1.15D-08	14	1.00D-28	6.16D-27	14	1.44D-07	2.54D-06	14	3.96D-16	2.82D-14
15	4.11D-10	5.58D-09	15	3.90D-30	2.37D-28	15	1.08D-07	2.32D-06	15	2.62D-16	2.50D-14
16	2.01D-10	2.75D-09	16	1.60D-31	9.64D-30	16	5.31D-08	9.51D-07	16	1.61D-17	1.12D-15
17	9.96D-11	1.38D-09	17	6.88D-33	4.11D-31	17	3.95D-08	8.66D-07	17	1.06D-17	9.86D-16
18	5.01D-11	6.98D-10	18	3.10D-34	1.84D-32	18	2.04D-08	3.70D-07	18	8.14D-19	4.92D-17
19	2.72D-11	3.59D-10	19	1.53D-35	8.59D-34	19	1.51D-08	3.35D-07	19	5.51D-19	4.28D-17
20	2.34D-11	1.86D-10	20	2.46D-36	4.17D-35	20	8.06D-09	1.49D-07	20	2.33D-19	2.35D-18
Table 6.1.1			Table 6.1.2			Table 6.1.3			Table 6.1.4		

Tables 6.1.1-6.1.4: $W_{80,k} = ||r(s_{80,k})||/||r(x_0)||$, where $s_{n,k}$ is computed by applying RRE in conjunction with the iterative method $x_{j+1} = Ax_j + b$, $j \geq 0$, A being given as in (6.1) with $N = 1000$. We take $b = 0$ so that the solution to $x = Ax + b$ is zero. The vector x_0 is picked as $(1, 1/\sqrt{2}, 1/\sqrt{3}, ..., 1/\sqrt{N})^T$. The $s_{n,k}$ and the corresponding upper bounds $\tilde{\Gamma}_{n,k}$ for $\Gamma^D_{n,k}$ are obtained by picking

(i) $\rho = \tau = \sigma/2$, $\sigma = 0.48$ so that $D = [0, 0.96]$ for Table 6.1.1,

(ii) $\rho = \tau = -\sigma/2$, $\sigma = -0.48$ so that $D = [-0.96, 0]$ for Table 6.1.2,

(iii) $\sigma = 0$, $\rho = \tau = 0.48$ so that $D = [-0.96, 0.96]$ for Table 6.1.3,

(iv) $\sigma = 0$, $\rho = -\tau = 0.48$ so that $D = \{\lambda = i\xi : -0.96 \leq \xi \leq 0.96\}$ for Table 6.1.4.

k	$W_{80,k}$	$\Gamma^{est}_{80,k}$
0	1.65D-02	2.75D-01
1	1.59D-02	2.71D-01
2	2.31D-03	7.46D-02
3	2.15D-03	7.28D-02
4	1.13D-03	2.70D-02
5	9.34D-04	2.62D-02
6	7.87D-04	1.13D-02
7	7.25D-04	1.09D-02
8	5.36D-04	5.16D-03
9	4.64D-04	4.95D-03
10	3.95D-04	2.51D-03
11	3.60D-04	2.40D-03
12	2.98D-04	1.28D-03
13	2.68D-04	1.22D-03
14	2.29D-04	6.73D-04
15	2.08D-04	6.39D-04
16	1.80D-04	3.65D-04
17	1.64D-04	3.46D-04
18	1.42D-04	2.03D-04
19	1.29D-04	1.92D-04
20	1.14D-04	1.15D-04

Table 6.1.5: $W_{80,k} = ||r(s_{80,k})||/||r(x_0)||$, where $s_{n,k}$ is computed by applying RRE in conjunction with the iterative method $x_{j+1} = Ax_j + b$, $j \geq 0$, A being given as in (6.1) with $N = 1000$ and $\sigma = 0$, $\rho = 0.6$, and $\tau = 0.384$. We take $b = 0$ so that the solution to $x = Ax + b$ is zero. The vector x_0 is picked as $(1, 1/\sqrt{2}, 1/\sqrt{3}, ..., 1/\sqrt{N})^T$. $\Gamma^{est}_{n,k}$ is obtained by letting $\beta = 0.984$ in (4.38), i.e., it is the corresponding $\tilde{\Gamma}_{n,k}$ appropriate for a mixed spectrum in $[-0.984, 0.984]$, although the spectrum of A is actually in $[-0.96, 0.96]$.

Figure Captions:

Figures 6.1.1-6.1.4: $\log_{10}(||r(x_i)||/||r(x_{init})||)$, $0 \le i \le n$, and $\log_{10}(||r(s_{n,k})||/||r(x_{init})||)$, $1 \le k \le K$, for (i) $n = 0$ and $K = 10$ and (ii) $n = 50$ and $K = 10$ (with the exception of Figure 6.1.2, for which $n = 20$ instead of $n = 50$), versus the cost of computing the x_i or the $s_{n,k}$ in the cycling mode. RRE is being applied in conjunction with (1.2), where A is as given in (6.1) with $N = 1000$, and $b = 0$ so that the solution is zero. Here x_{init} is the initial vector given as $x_{init} = (1, 1/\sqrt{2}, 1/\sqrt{3}, ..., 1/\sqrt{N})^T$.

(i) $\rho = \tau = \sigma/2$, $\sigma = 0.48$ for Figure 6.1.1,
(ii) $\rho = \tau = -\sigma/2$, $\sigma = -0.48$ for Figure 6.1.2,
(iii) $\sigma = 0$, $\rho = \tau = 0.48$ for Figure 6.1.3,
(iv) $\sigma = 0$, $\rho = -\tau = 0.48$ for Figure 6.1.4.

Figure 6.1.5: $\log_{10}(||r(x_i)||/||r(x_{init})||)$, $0 \le i \le n$, and $\log_{10}(||r(s_{n,k})||/||r(x_{init})||)$, $1 \le k \le K$, for (i) $n = 0$ and $K = 10$ and (ii) $n = 50$ and $K = 10$, versus the cost of computing the x_i or the $s_{n,k}$ in the cycling mode. RRE is being applied in conjunction with (1.2), where A is given in (6.1) with $\sigma = 0$, $\rho = 0.6$, and $\tau = 0.384$, and $N = 1000$, and $b = 0$ so that the solution is zero. Here x_{init} is the initial vector given as $x_{init}(1, 1/\sqrt{2}, 1/\sqrt{3}, ..., 1/\sqrt{N})^T$.

Figures 6.2.1-6.2.3: $\log_{10}||r(x_i)||$, $0 \le i \le n$, and $\log_{10}||r(s_{n,k})||$, $1 \le k \le K$, for (i) $n = 0$ and $K = 20$ and (ii) $n = 50$ and $K = 20$, versus the cost of computing the x_i or the $s_{n,k}$ in the cycling mode. RRE is applied in conjunction with the Jacobi iteration to the linear system arising from the discretization of the convection-diffusion equation in Example 6.2. Here the solution is zero and $x_{init} = (1, 1/\sqrt{2}, 1/\sqrt{3}, ..., 1/\sqrt{N})^T$ is the initial vector.

(i) $\gamma = 100$, $\beta = 0$ for Figure 6.2.1,
(ii) $\gamma = 100$, $\beta = -100$ for Figure 6.2.2,
(iii) $\gamma = 125$, $\beta = -100$ for Figure 6.2.3.

Figures 6.2.4-6.2.6: $\log_{10}||r(x_i)||, 0 \leq i \leq n$, and $\log_{10}||r(s_{n,k})||, 1 \leq k \leq K$, for (i) $n = 0$ and $K = 10$ and (ii) $n = 25$ and $K = 10$, versus the cost of computing the x_i or the $s_{n,k}$ in the cycling mode. RRE is applied in conjunction with the double Jacobi iteration to the linear system arising from the discretization of the convection-diffusion equation in Example 6.2. Here the solution is zero and $x_{init} = (1, 1/\sqrt{2}, 1/\sqrt{3}, ..., 1/\sqrt{N})^T$ is the initial vector.

(i) $\gamma = 100$, $\beta = 0$ for Figure 6.2.4,

(ii) $\gamma = 100$, $\beta = -100$ for Figure 6.2.5,

(iii) $\gamma = 125$, $\beta = -100$ for Figure 6.2.6.

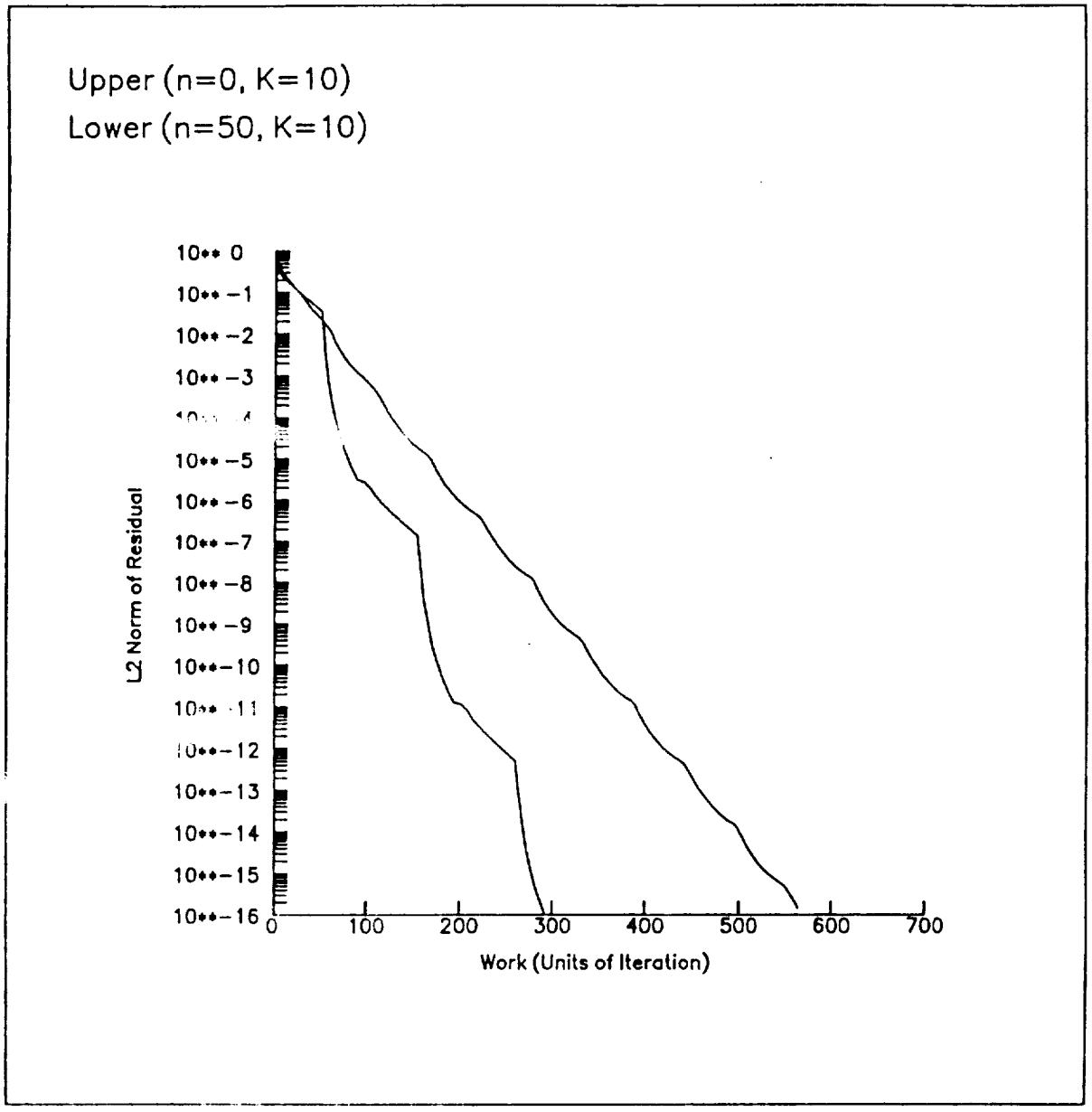
Upper (n=0, K=10)
Lower (n=50, K=10)
10** 0
10** -1
10** -2
10** -3
10** -5
10** -6
10** -7
10** -8
10** -9
10**-10
10**-12
10**-13
10**-14
10**-15
10**-16
L2 Norm of Residual
0
100
200
300
400
500
600
700
Work (Units of Iteration)

Fig. 6.1.1

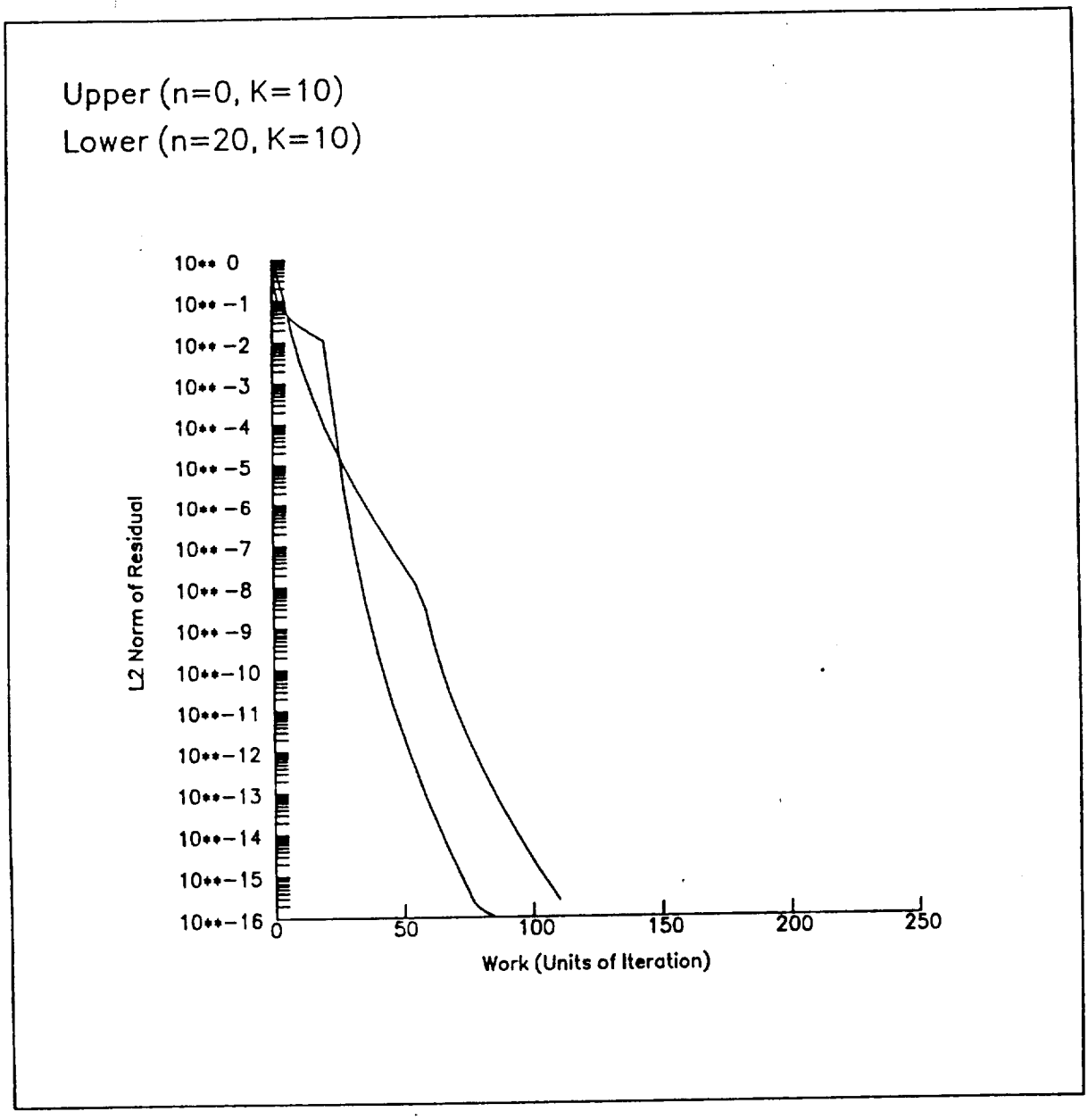
Upper (n=0, K=10)
Lower (n=20, K=10)
L2 Norm of Residual
10** 0
10** -1
10** -2
10** -3
10** -4
10** -5
10** -6
10** -7
10** -8
10** -9
10**-10
10**-11
10**-12
10**-13
10**-14
10**-15
10**-16
0
50
100
150
200
250
Work (Units of Iteration)

Fig. 6.1.2

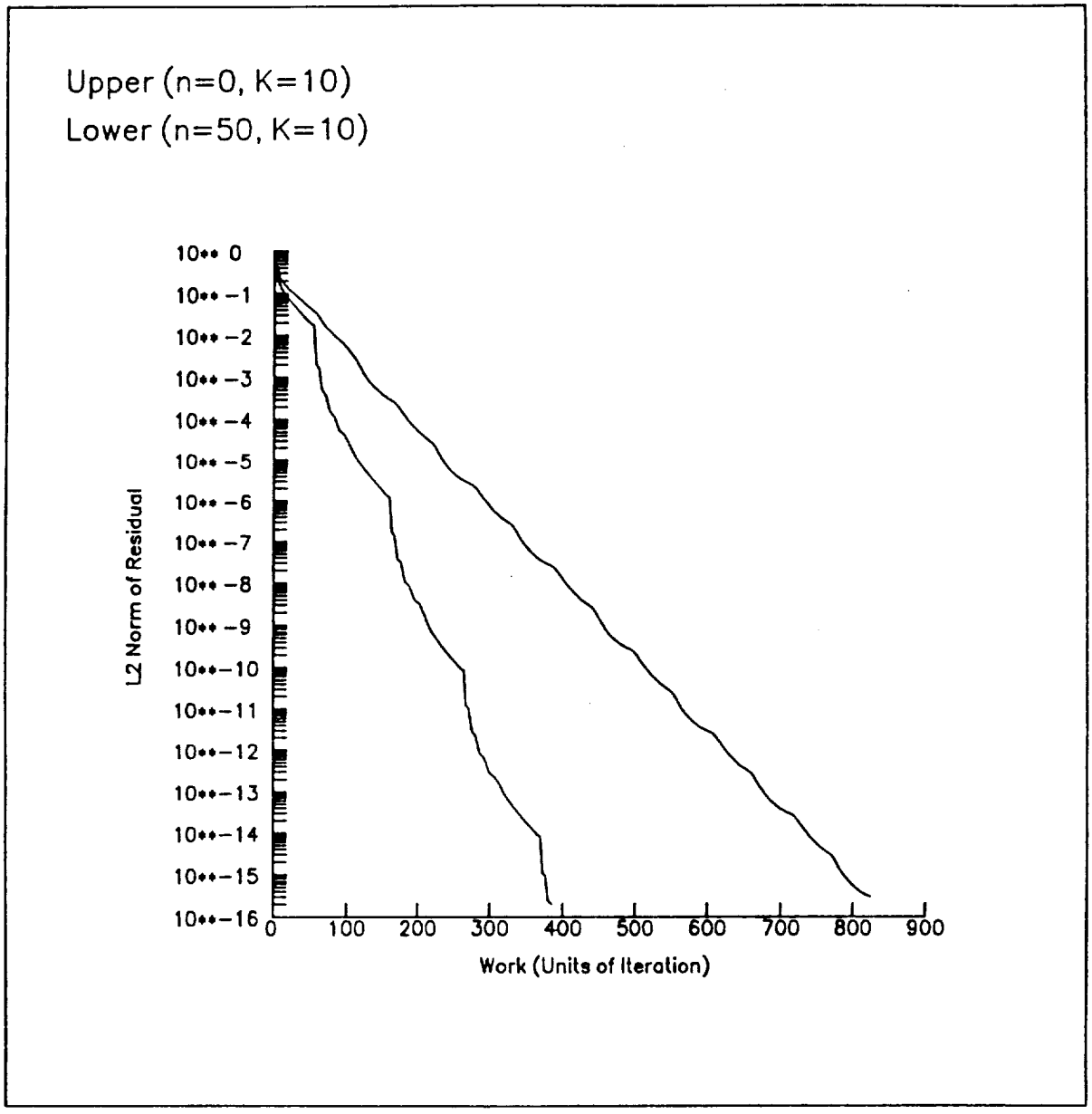
Upper (n=0, K=10)
Lower (n=50, K=10)
10** 0
10** -1
10** -2
10** -3
10** -4
10** -5
10** -6
10** -7
10** -8
10** -9
10**-10
10**-11
10**-12
10**-13
10**-14
10**-15
10**-16
L2 Norm of Residual
0
100
200
300
400
500
600
700
800
900
Work (Units of Iteration)

Fig. 6.1.3

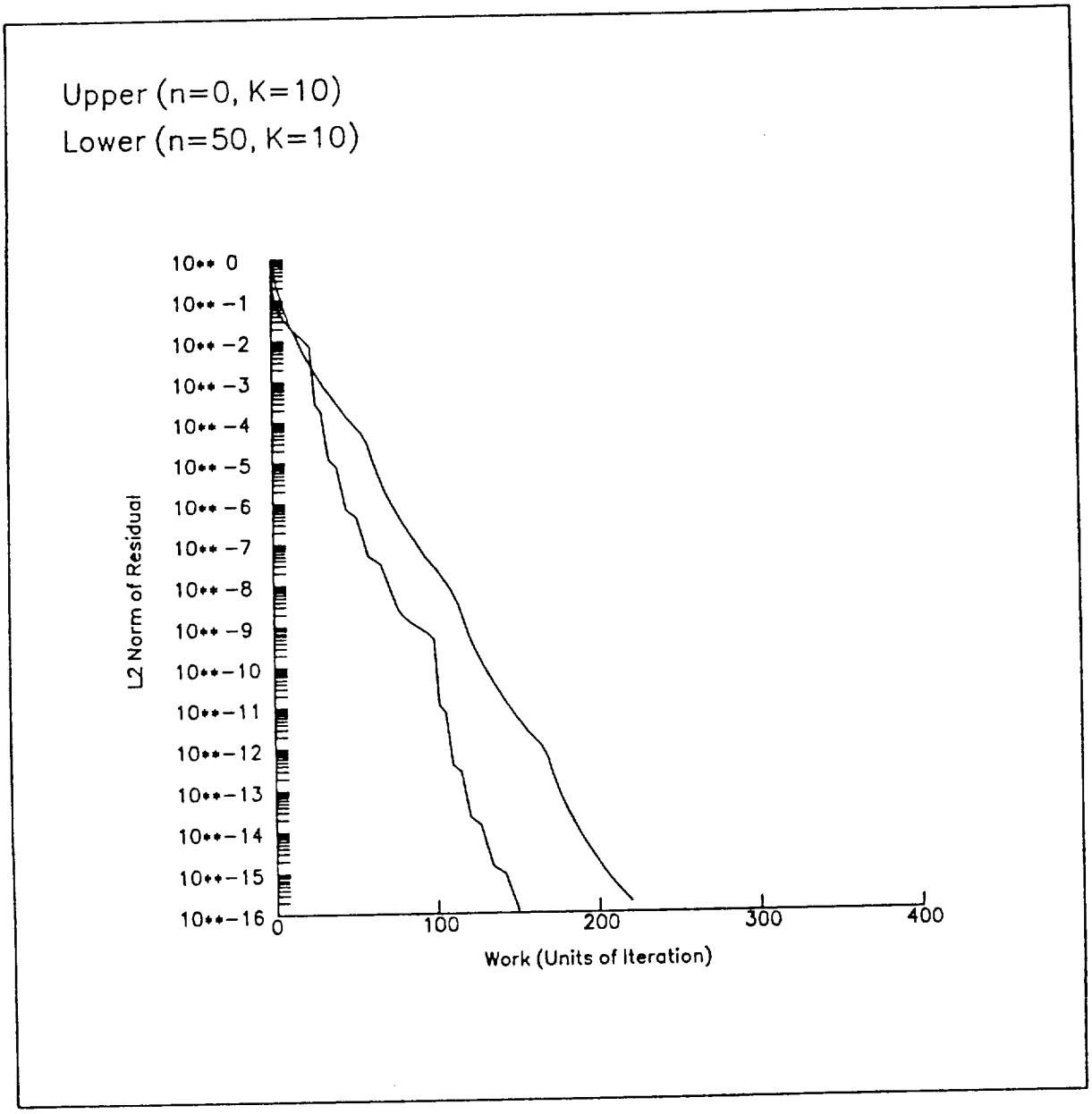
Upper (n=0, K=10)
Lower (n=50, K=10)
10** 0
10** -1
10** -2
10** -3
10** -4
10** -5
10** -6
10** -7
10** -8
10** -9
10**-10
10**-11
10**-12
10**-13
10**-14
10**-15
10**-16
L2 Norm of Residual
0
100
200
300
400
Work (Units of Iteration)

Fig. 6.1.4

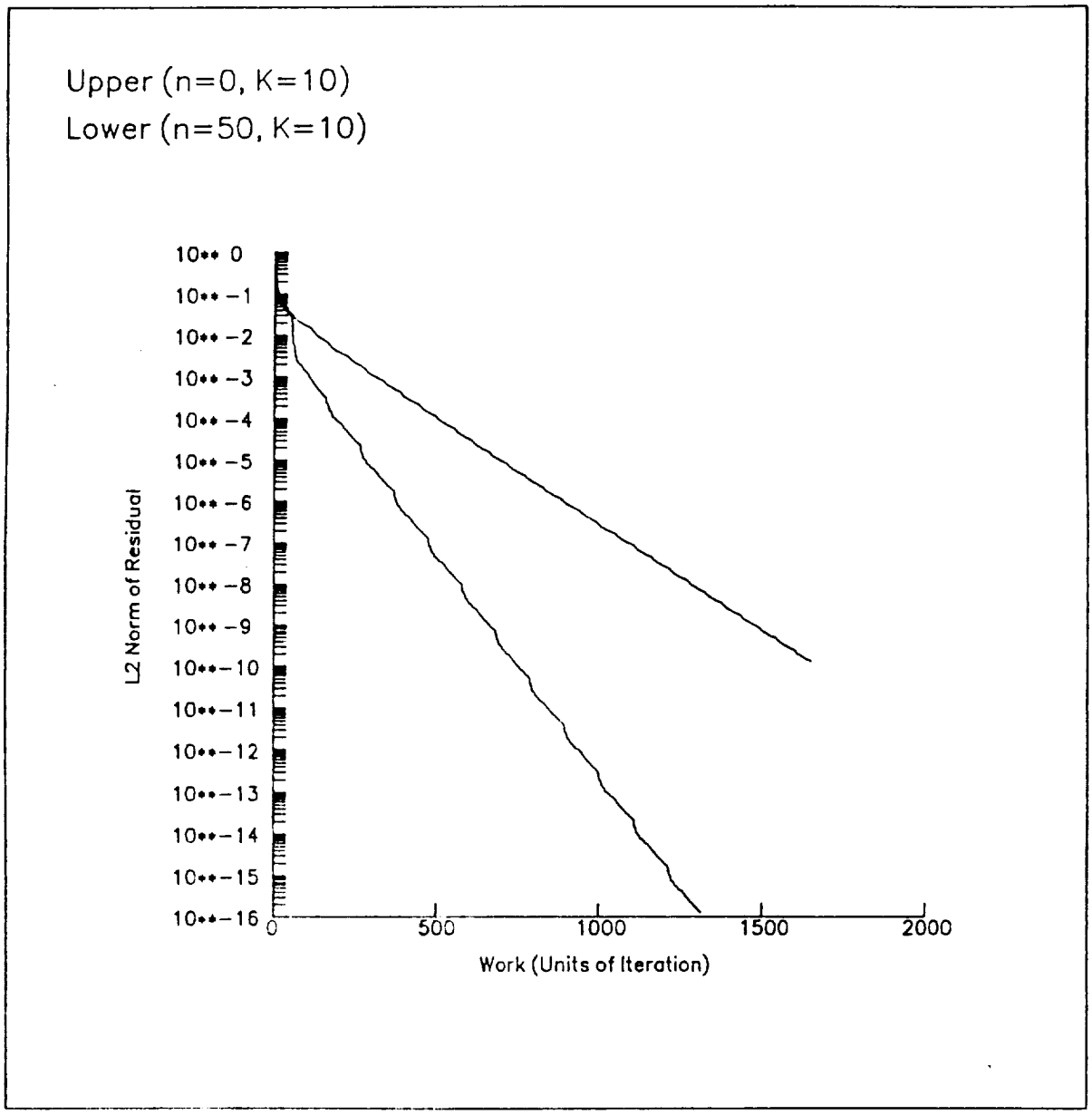
Upper (n=0, K=10)
Lower (n=50, K=10)
10** 0
10** -1
10** -2
10** -3
10** -4
10** -5
10** -6
10** -7
10** -8
10** -9
10**-10
10**-11
10**-12
10**-13
10**-14
10**-15
10**-16
L2 Norm of Residual
0
500
1000
1500
2000
Work (Units of Iteration)

Fig. 6.1.5

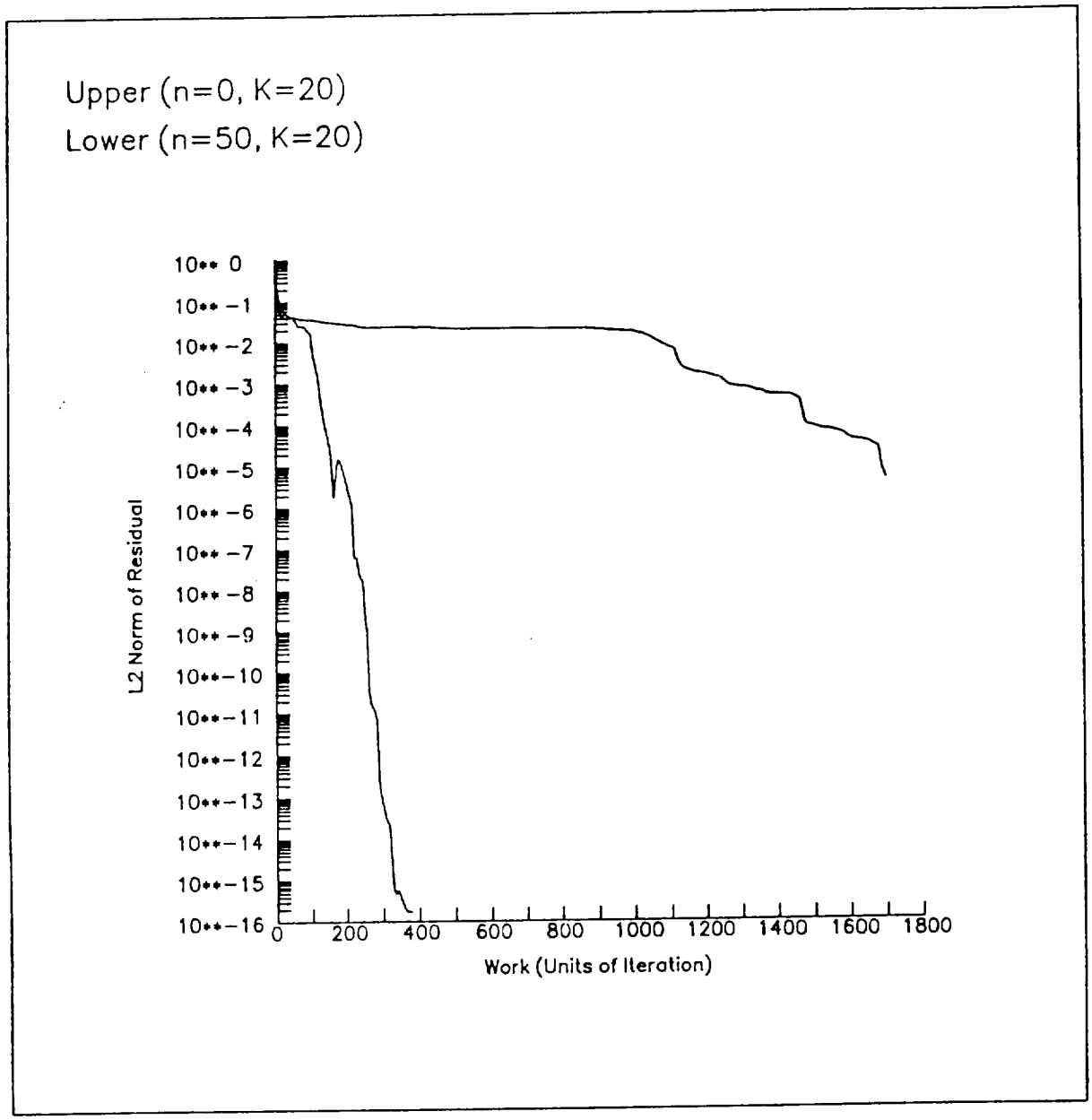
Upper (n=0, K=20)
Lower (n=50, K=20)
L2 Norm of Residual
10** 0
10** -1
10** -2
10** -3
10** -4
10** -5
10** -6
10** -7
10** -8
10** -9
10**-10
10**-11
10**-12
10**-13
10**-14
10**-15
10**-16
0
200
400
600
800
1000
1200
1400
1600
1800
Work (Units of Iteration)

Fig. 6.2.1

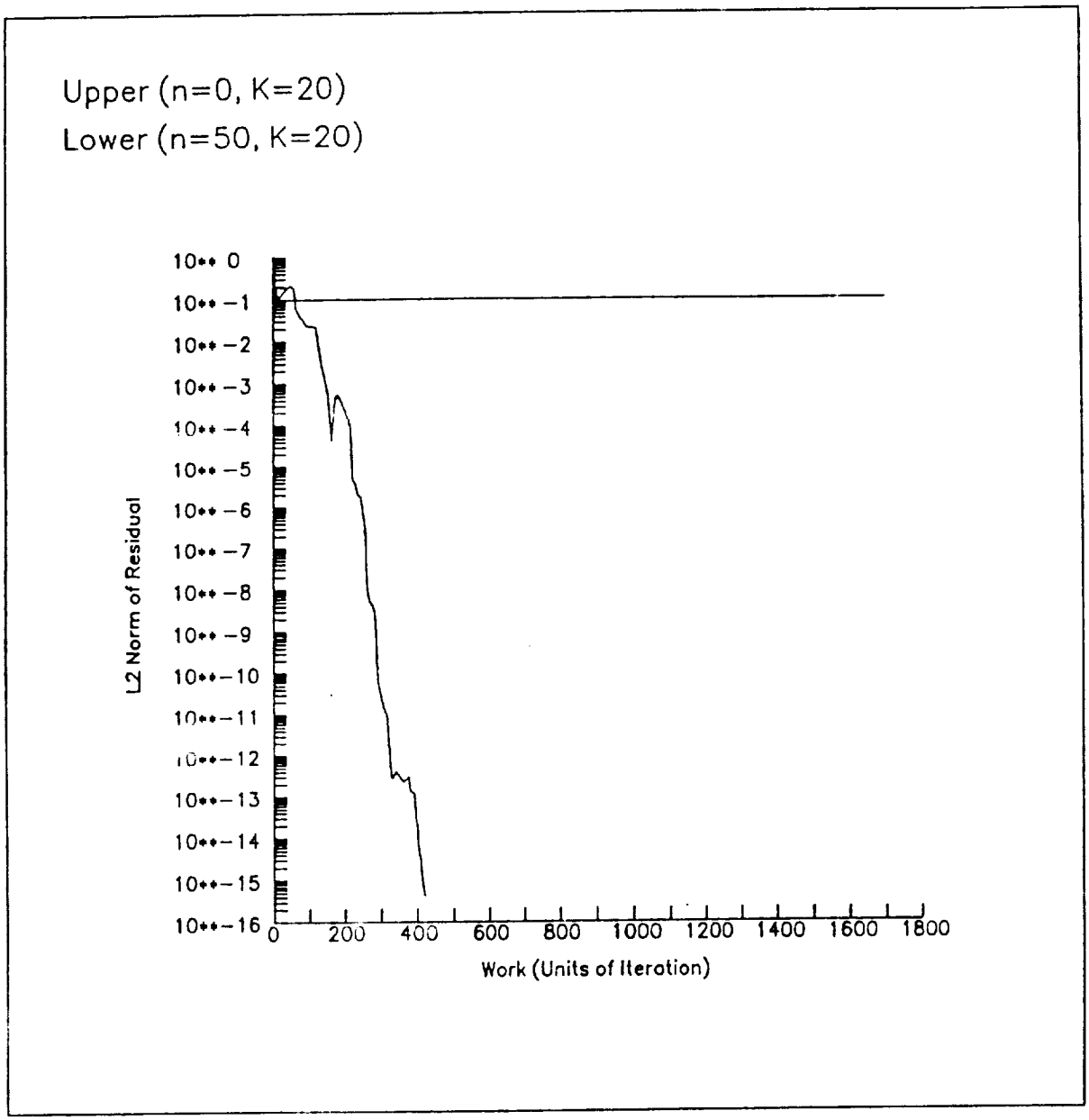
Upper (n=0, K=20)
Lower (n=50, K=20)
L2 Norm of Residual
10** 0
10**-1
10**-2
10**-3
10**-4
10**-5
10**-6
10**-7
10**-8
10**-9
10**-10
10**-11
10**-12
10**-13
10**-14
10**-15
10**-16
0
200
400
600
800
1000
1200
1400
1600
1800
Work (Units of Iteration)

Fig. 6.2.2

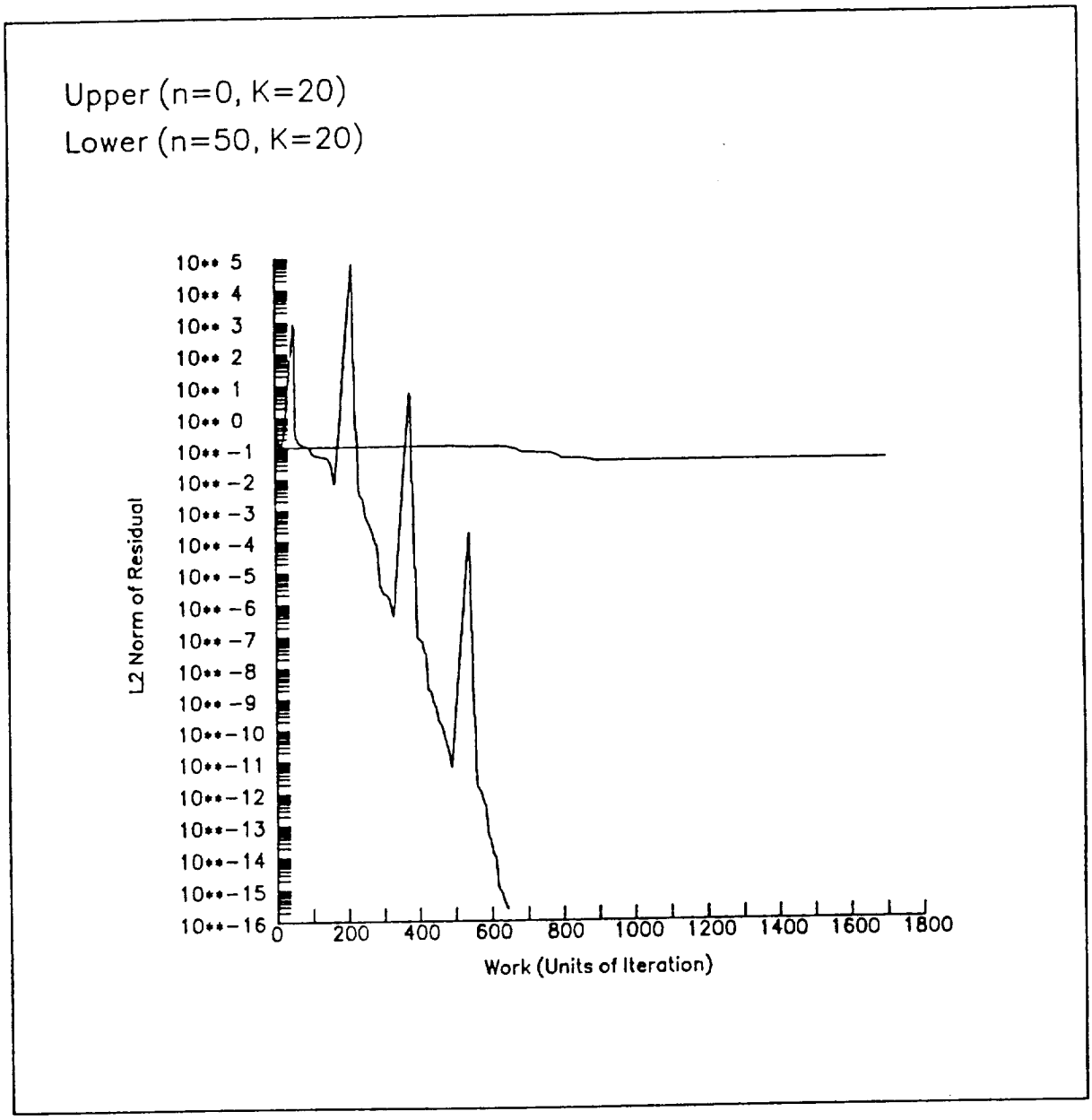
Upper (n=0, K=20)
Lower (n=50, K=20)
L2 Norm of Residual
10** 5
10** 4
10** 3
10** 2
10** 1
10** 0
10** -1
10** -2
10** -3
10** -4
10** -5
10** -6
10** -7
10** -8
10** -9
10**-10
10**-11
10**-12
10**-13
10**-14
10**-15
10**-16
0
200
400
600
800
1000
1200
1400
1600
1800
Work (Units of Iteration)

Fig. 6.2.3

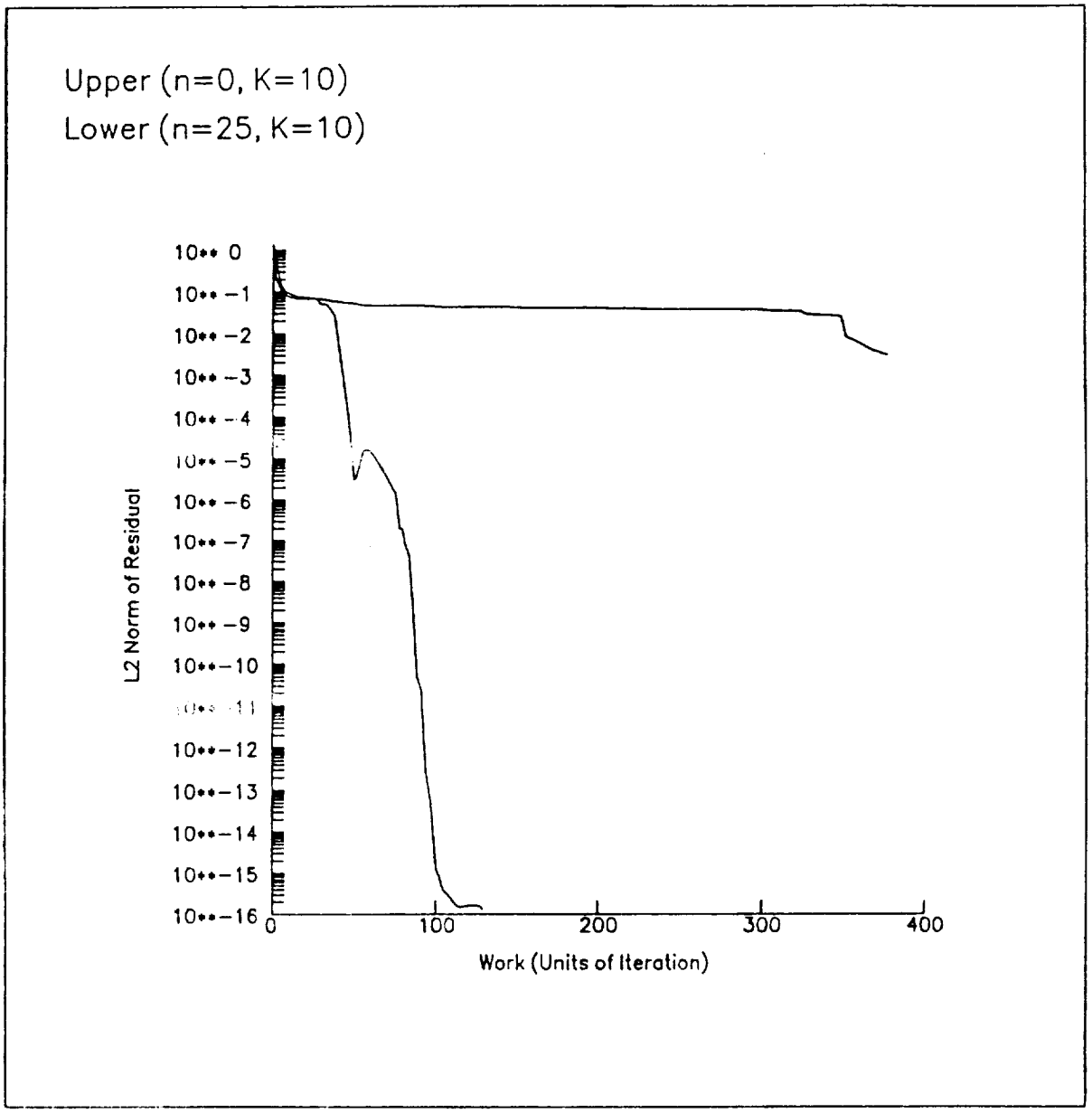
Upper (n=0, K=10)
Lower (n=25, K=10)
L2 Norm of Residual
10** 0
10** -1
10** -2
10** -3
10** -4
10** -5
10** -6
10** -7
10** -8
10** -9
10**-10
10**-12
10**-13
10**-14
10**-15
10**-16
0
100
200
300
400
Work (Units of Iteration)

Fig. 6.2.4

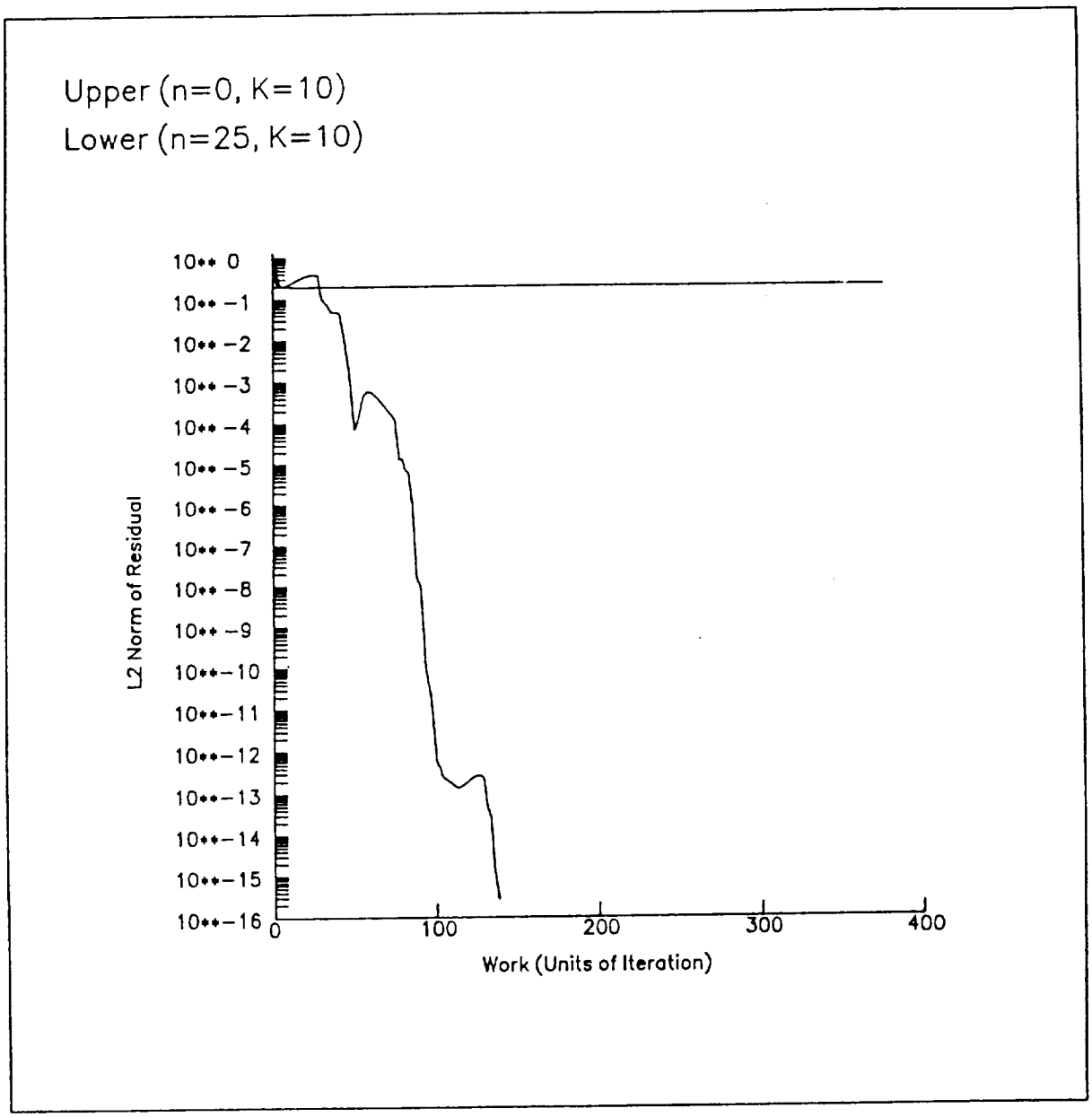
Upper (n=0, K=10)
Lower (n=25, K=10)
L2 Norm of Residual
10** 0
10** -1
10** -2
10** -3
10** -4
10** -5
10** -6
10** -7
10** -8
10** -9
10**-10
10**-11
10**-12
10**-13
10**-14
10**-15
10**-16
0
100
200
300
400
Work (Units of Iteration)

Fig. 6.2.5

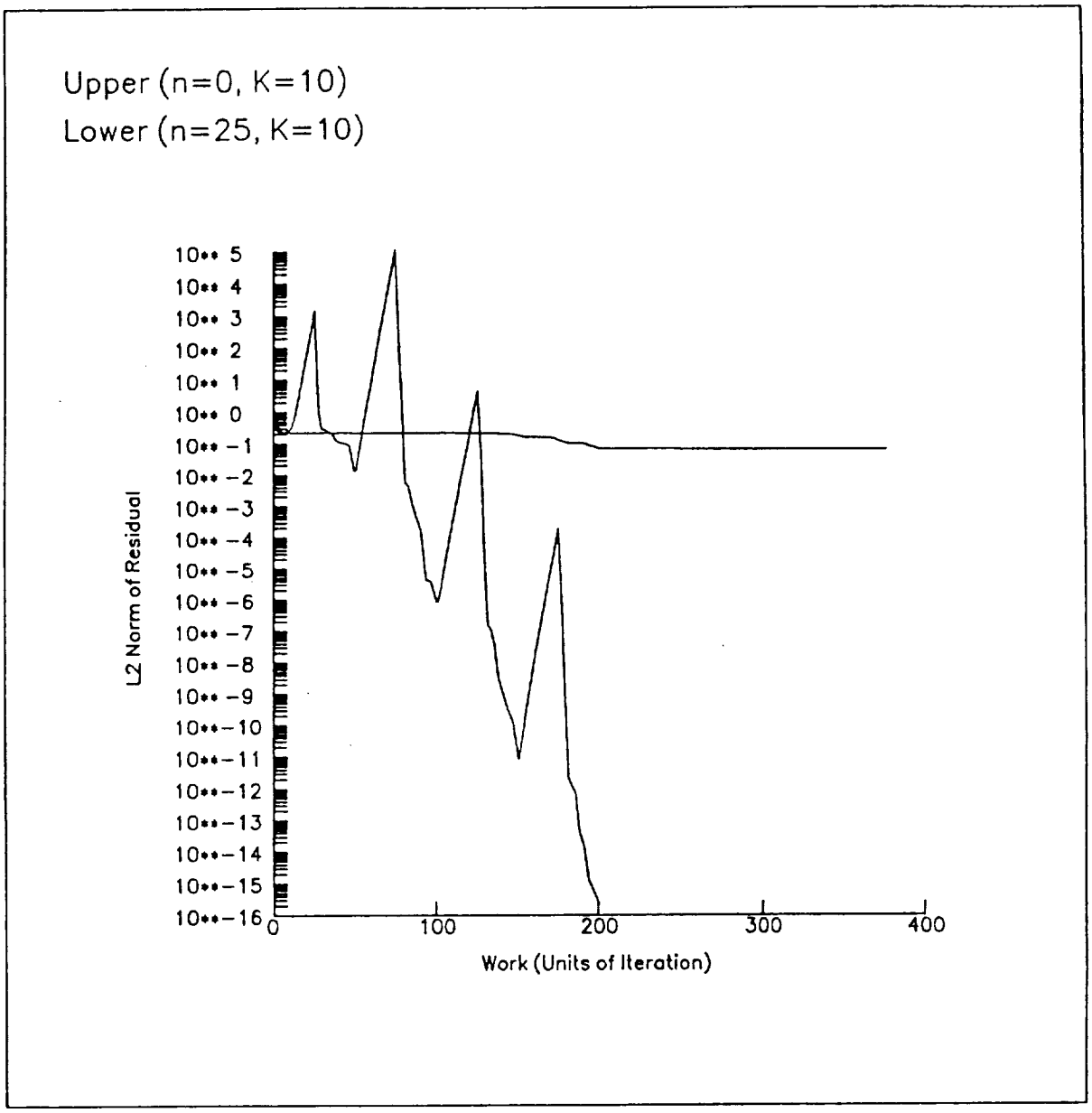
Upper (n=0, K=10)
Lower (n=25, K=10)
L2 Norm of Residual
10** 5
10** 4
10** 3
10** 2
10** 1
10** 0
10** -1
10** -2
10** -3
10** -4
10** -5
10** -6
10** -7
10** -8
10** -9
10**-10
10**-11
10**-12
10**-13
10**-14
10**-15
10**-16
0
100
200
300
400
Work (Units of Iteration)

Fig. 6.2.6

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. AGENCY USE ONLY (Leave blank)	2. REPORT DATE	3. REPORT TYPE AND DATES COVERED
	1992	Technical Memorandum

4. TITLE AND SUBTITLE

Upper Bounds for Convergence Rates of Vector Extrapolation Methods on Linear Systems With Initial Iterations

5. FUNDING NUMBERS

WU-505-62-21

6. AUTHOR(S)

Avram Sidi and Yair Shapira

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)

National Aeronautics and Space Administration
Lewis Research Center
Cleveland, Ohio 44135-3191

8. PERFORMING ORGANIZATION REPORT NUMBER

E-6946

9. SPONSORING/MONITORING AGENCY NAMES(S) AND ADDRESS(ES)

National Aeronautics and Space Administration
Washington, D.C. 20546-0001

10. SPONSORING/MONITORING AGENCY REPORT NUMBER

NASA TM-105608
ICOMP-92-09

11. SUPPLEMENTARY NOTES

Avram Sidi, Computer Science Department, Technion-Israel Institute of Technology, Haifa 32000, Israel, and Institute for Computational Mechanics in Propulsion, NASA Lewis Research Center, Cleveland, Ohio 44135; Yair Shapira, Mathematics Department, Technion-Israel Institute of Technology, Haifa 32000, Israel. This report was submitted by Yair Shapira as a thesis in partial fulfillment of the requirements for the degree Doctor of Science at the Technion-Israel Institute of Technology. Project Manager, L.A. Povinelli, Lewis Research Academy, NASA Lewis Research Center, (216) 433-5818.

12a. DISTRIBUTION/AVAILABILITY STATEMENT

Unclassified - Unlimited
Subject Category 64

12b. DISTRIBUTION CODE

13. ABSTRACT *(Maximum 200 words)*

The application of the minimal polynomial extrapolation (MPE) and the reduced rank extrapolation (RRE) to a vector sequence obtained by the linear iterative technique $x_{j+1} = Ax_j + b$, $j = 1, 2,...$, is considered. Both methods produce a two-dimensional array of approximations $s_{n,k}$ to the solution of the system $(I - A)x = b$. Here $s_{n,k}$ is obtained from the vectors x_j, $n \le j \le n + k + 1$. It was observed in an earlier publication by the first author that the sequence $s_{n,k}$, $k = 1, 2,...$, for $n > 0$, but fixed, possesses better convergence properties that the sequence $s_{0,k}$, $k = 1, 2,...$. A detailed theoretical explanation for this phenomenon is provided in the present work. This explanation is heavily based on approximations by incomplete polynomials. It is demonstrated by numerical examples when the matrix A is sparse that cycling with $s_{n,k}$ for $n > 0$, but fixed, produces better convergence rates and costs less computationally than cycling with $s_{0,k}$. It is also illustrated numerically with a convection-diffusion problem that the former may produce excellent results where the latter may fail completely. As has been shown in an earlier publication, the results produced by $s_{0,k}$ are identical to the corresponding results obtained by applying the Arnoldi method or GMRES to the system $(I - A)x = b$.

14. SUBJECT TERMS	15. NUMBER OF PAGES
Minimal polynomial extrapolation; Reduced rank extrapolation; Vector extrapolation methods	58
	16. PRICE CODE: A04

17. SECURITY CLASSIFICATION OF REPORT	18. SECURITY CLASSIFICATION OF THIS PAGE	19. SECURITY CLASSIFICATION OF ABSTRACT	20. LIMITATION OF ABSTRACT
Unclassified	Unclassified	Unclassified	